

NUMERICAL METHODS FOR A CLASS OF NONLOCAL DIFFUSION PROBLEMS WITH THE USE OF BACKWARD SDES *

G. ZHANG[†], W. ZHAO[‡], C. G. WEBSTER[†], AND M. GUNZBURGER[§]

Abstract. We propose a novel numerical approach for nonlocal diffusion equations [8] with integrable kernels, based on the relationship between the backward Kolmogorov equation and backward stochastic differential equations (BSDEs) driven by Lévy processes with jumps. The nonlocal diffusion problem under consideration is converted to a BSDE, for which numerical schemes are developed and applied directly. As a stochastic approach, the proposed method does not require the solution of linear systems, which allows for embarrassingly parallel implementations and also enables adaptive approximation techniques to be incorporated in a straightforward fashion. Moreover, our method is more accurate than classic stochastic approaches due to the use of high-order temporal and spatial discretization schemes. In addition, our approach can handle a broad class of problems with general nonlinear forcing terms as long as they are globally Lipschitz continuous. Rigorous error analysis of the new method is provided as several numerical examples that illustrate the effectiveness and efficiency of the proposed approach.

Key words. backward stochastic differential equation with jumps, nonlocal diffusion equations, superdiffusion, compound Poisson process, θ -scheme, adaptive approximation

1. Introduction. A diffusion process is deemed nonlocal when the associated underlying Lévy process does not only consist of Brownian motions. The features of nonlocal diffusion has been verified experimentally to be present in a wide variety of applications, including, e.g., contaminant flow in groundwater, plasma physics, and the dynamics of financial markets. A comprehensive survey of nonlocal diffusion problems is given in [15]. In this work, we consider a partial-integral diffusion equation (PIDE) representation [6, 11, 8] of linear nonlocal diffusion problems. Since it is typically difficult to obtain analytical solutions of such problems, numerical solutions are highly desired in practical applications. There are mainly two types of numerical methods for nonlocal diffusion equations under consideration. The first is the family of deterministic approaches, e.g., finite-element-type methods [6, 11] and collocation methods, whereas the second can be classified as stochastic approaches, e.g., continuous-time random walk (CTRW) methods [5, 10, 15].

The deterministic approaches are extensions of existing methods for local partial differential equations (PDEs) that incorporate schemes to discretize the underlying integral operators. The recently developed nonlocal vector calculus [9] provides helpful tools that allow one to analyze nonlocal diffusion problems in a similar manner to analyzing local PDEs; see [8, 11] for details. However, in the context of implicit time-stepping schemes, the nonlocal operator may result in severe computational difficulties coming from the dramatic deterioration of the sparsity of the stiffness matrices

*This material is based upon work supported in part by the U.S. Air Force of Scientific Research under grant numbers FA9550-11-1-0149 and 1854-V521-12; by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research, Applied Mathematics program under contract numbers ERKJ259, and ERKJE45; by the National Natural Science Foundations of China under grant numbers 91130003 and 11171189; by Natural Science Foundation of Shandong Province under grant number ZR2011AZ002; and by the Laboratory Directed Research and Development program at the Oak Ridge National Laboratory, which is operated by UT-Battelle, LLC, for the U.S. Department of Energy under Contract DE-AC05-00OR22725.

[†]Department of Computational and Applied Mathematics, Oak Ridge National Laboratory, Oak Ridge, TN 37831(zhangg@ornl.gov, webstercg@ornl.gov).

[‡]School of Mathematics, Shandong University, Jinan 250100, China (wdzhao@sdu.edu.cn).

[§]Department of Scientific Computing, Florida State University (gunzburg@fsu.edu).

required by the underlying linear systems. On the other hand, stochastic approaches, e.g., CTRW methods, are based on the relation between nonlocal diffusion and a class of Lévy jump processes. Although stochastic methods do not require the solution of linear systems, and the simulations of all paths of the random walk can be easily parallelized, they also have some drawbacks. Typically, most stochastic methods are sampling-based Monte Carlo approaches, suffering from slow convergence, thus requiring a very large number of samples to achieve small errors. Even worse, in the case of a general nonlinear forcing term, the nonlocal diffusion equation is no longer the master equation of the underlying jump processes.

In this paper, we propose a novel stochastic numerical scheme for the linear nonlocal diffusion problems studied in [5, 6, 10, 9, 11, 8] based on the relationship between the PIDEs and a certain class of backward stochastic differential equations (BSDEs) with jumps. The existence and uniqueness of solutions for nonlinear BSDEs with and without jumps have been proved in [17] and [1], respectively. Since then, BSDEs have become important tools in probability theory, stochastic optimal control and mathematical finance [7]. Unlike BSDEs driven by Brownian motions, there are very few numerical schemes proposed for BSDEs with jumps, and most of the schemes are focussed on time discretizations only. For example, a numerical scheme based on Picard's method was proposed in [14], and a forward Euler scheme was proposed in [2, 3] where the convergence rate was proved to be $(\Delta t)^{\frac{1}{2}}$.

Our focus will be on the BSDEs corresponding to nonlocal diffusion equations with *integrable* jump kernels in unbounded domains. Since the PIDEs of interest do not possess local convection and diffusion terms, the corresponding BSDEs can be simplified, such that, the underlying Lévy processes are merely *compound Poisson processes*. Also, motivated by various engineering applications discussed in [8], our goal is to develop and analyze computationally efficient numerical schemes that can achieve high-order accuracy in both temporal and spatial discretization, rather than focus on the scalability of our schemes in solving extremely high-dimensional problems (e.g., in option pricing). From this perspective, although the BSDEs under consideration is a special case of the equations studied in [2], high-order fully-discrete schemes and the corresponding error analysis are still missing in the literature. In fact, based on our previous work, it is particularly challenging to recover the classic finite element convergence rates in the spatial discretization, even for the simplified BSDEs. Thus, the main contributions of this paper are summarized as follows:

- Development of stable high-order matrix-free numerical schemes for the BSDEs and the corresponding nonlocal diffusion problems;
- Analysis of the approximation error of the proposed semi-discrete and fully-discrete schemes, and;
- Numerical demonstration of the capabilities of handling multiple types of integrable kernels, e.g., symmetric, non-symmetric, or singular kernels.

Specifically, the BSDEs will be discretized, in the temporal domain, using a θ -scheme extended from our previous works [21, 20] for BSDEs without jumps. In particular, the cases where $\theta = 0$, $\theta = \frac{1}{2}$ and $\theta = 1$ correspond to forward Euler, Crank-Nicolson and backward Euler schemes, respectively. As discussed in [21], a quadrature rule adapted to the underlying stochastic processes is critical to approximate all the conditional mathematical expectations in our numerical scheme. Thus, we propose a general formulation of the quadrature rule for estimating the conditional expectations with respect to the compound Poisson processes, and a specific form of a certain PIDE can be determined based on regularities of the kernel and forcing

term. In §5, Gauss-Legendre, Gauss-Jacobi and Newton-Cotes rules are substituted into the proposed quadrature rule to approximate different kernels and forcing terms. In addition, since the total number of quadrature points grows exponentially with the number of time steps, we construct a piecewise Lagrange interpolation scheme based on a pre-determined spatial mesh, in order to evaluate the integrand of the expectations at all quadrature points.

Both theoretical analysis and numerical experiments show that our approach is advantageous for linear nonlocal diffusion equations with integrable kernels on unbounded domains. Compared to deterministic approaches, e.g., finite elements and collocation approaches, *the proposed methods do not require the solution of dense linear systems*. Instead, the PIDE can be solved independently at different spatial grid points on each time level, making it straightforward to incorporate parallel implementation and adaptive spatial approximation. Compared to stochastic approaches, our scheme is more accurate than the CTRW method due to the use of the θ -scheme for time discretization, the high-order quadrature rule, and piecewise Lagrange polynomial interpolation for spatial discretization. In addition, our method can handle a broad class of problems with general nonlinear forcing terms as long as they are globally Lipschitz continuous.

The outline of the paper is as follows. In §2, we introduce the mathematical description of the linear nonlocal diffusion equations and the corresponding class of BSDEs under consideration. In §3, we propose our numerical schemes for the BSDEs of interest, where the semi-discrete and the fully-discrete approximations are presented in §3.1 and §3.2, respectively. Error estimates for the proposed scheme are proved in §4. Extensive numerical examples and comparisons to existing techniques are given in §5, which are shown to be consistent with the theoretical results and reveal the effectiveness of our approach. Finally, several concluding remarks are given in §6.

2. Nonlocal diffusion models and BSDEs with jumps. This section is dedicated to describing the nonlocal diffusion problem that is the focus of this paper. In particular, in §2.1 we introduce the definitions of a general nonlocal operator and diffusion equation. The relationship between the backward Kolmogorov equation, which is a generalization of the nonlocal diffusion equation of interest, and the BSDE with jumps, is described in §2.2. Based on this equivalence, a simplified BSDE corresponding to the nonlocal diffusion problem of §2.1 is also given at the end of §2.2.

2.1. Nonlocal diffusion equations. Let us recall a nonlocal operator introduced in [8]. For a function $u(t, \mathbf{x}) : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$, with $d = 1, 2, 3$ and $T > 0$, we define the action of the linear operator \mathcal{L} on $u(t, \mathbf{x})$ as

$$\mathcal{L}u = \int_{\mathbb{R}^d} (u(t, \mathbf{x} + \mathbf{e}) - u(t, \mathbf{x})) \gamma(\mathbf{e}) d\mathbf{e}, \quad \forall (t, \mathbf{x}) \in [0, T] \times \mathbb{R}^d, \quad (2.1)$$

where the properties of \mathcal{L} depend crucially on the kernel function $\gamma(\mathbf{e}) : \mathbb{R}^d \rightarrow \mathbb{R}$. In this work, we focus on a particular class of kernel functions which are nonnegative and integrable, i.e.,

$$\gamma(\mathbf{e}) \geq 0 \quad \forall \mathbf{e} \in \mathbb{R}^d \quad \text{and} \quad \int_{\mathbb{R}^d} \gamma(\mathbf{e}) d\mathbf{e} < \infty. \quad (2.2)$$

Note that $\gamma(\mathbf{e})$ may be *symmetric*, i.e., $\gamma(\mathbf{e}) = \gamma(-\mathbf{e})$ for any $\mathbf{e} \in \mathbb{R}^d$, or *non-symmetric*, i.e., there exists $\mathbf{e} \in \mathbb{R}^d$ such that $\gamma(\mathbf{e}) \neq \gamma(-\mathbf{e})$. The operator \mathcal{L} exhibits nonlocal behavior because, for a fixed $t \in [0, T]$, the value of $\mathcal{L}u$ at a point $\mathbf{x} =$

$(x^1, \dots, x^d) \in \mathbb{R}^d$ requires information about u at points $\mathbf{x} + \mathbf{e} \neq \mathbf{x}$; this is contrasted with local operators, e.g., the value of Δu at a point \mathbf{x} requires information about u only in an infinitesimal neighborhood of \mathbf{x} . Our interest in the operator \mathcal{L} is due to its participation in the *nonlocal diffusion equation*

$$\begin{cases} \frac{\partial u}{\partial t}(t, \mathbf{x}) - \mathcal{L}u(t, \mathbf{x}) = g(t, \mathbf{x}, u) & \text{for } (t, \mathbf{x}) \in (0, T] \times \mathbb{R}^d \\ u(0, \mathbf{x}) = u_0(\mathbf{x}), & \text{for } \mathbf{x} \in \mathbb{R}^d, \end{cases} \quad (2.3)$$

where $u_0(\mathbf{x})$ is the initial condition and the forcing term $g(t, \mathbf{x}, u)$ may be a *nonlinear* function of t, \mathbf{x} and u . We remark that (2.3) is a nonlocal Cauchy problem defined on the *unbounded* spatial domain \mathbb{R}^d . Also, in the context of kernel functions $\gamma(\mathbf{e})$ that are compactly supported in \mathbb{R}^d , the initial-boundary value nonlocal problem with *volume constraints*, i.e., constraints applied on a regions of nonzero measure in \mathbb{R}^d , has been well studied [6, 8, 11]. However, the well-posedness of the corresponding BSDEs with volume constraints is still an open challenge. As such, we do not impose volume constraints to (2.3) in this effort.

2.2. BSDEs and backward Kolmogorov equations. Now we discuss the probabilistic representation of the solution of the nonlocal diffusion equation in (2.3). The nonlinear Feynman-Kac formula studied in [16] shows that BSDEs driven by Brownian motion provides a probabilistic representation of the solutions of a class of second-order quasi-linear parabolic PDEs. This result was extended in [1] to the case of BSDEs driven by general Lévy processes with jumps. Due to the inclusion of Lévy jumps, the counterpart of a BSDE is a partial integral differential equation (PIDE), i.e., the backward Kolmogorov equation [13]. It turns out that such a PIDE is a generalization of the nonlocal diffusion equation in (2.3). Here we first introduce a general form of a BSDE with jumps and the backward Kolmogorov equation, then give a simplified form of the BSDE corresponding to the nonlocal diffusion equation (2.3) under consideration.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space with a filtration $\{\mathcal{F}_t\}_{0 \leq t \leq T}$ for a finite terminal time $T > 0$. We assume the filtration $\{\mathcal{F}_t\}_{0 \leq t \leq T}$ satisfies the usual hypotheses of completeness, i.e., \mathcal{F}_0 contains all sets of \mathbb{P} -measure zero and maintains right continuity, i.e., $\mathcal{F}_t = \mathcal{F}_{t+}$. Moreover, the filtration is assumed to be generated by two mutually independent processes, i.e., one m -dimensional Brownian motion $\mathbf{W}_t = (W_t^1, \dots, W_t^m)^\top$ and one d -dimensional Poisson random measure $\mu(A, t)$ on $E \times [0, T]$, where $E = \mathbb{R}^d \setminus \{0\}$ is equipped with its Borel field \mathcal{E} . The compensator of μ and the resulting compensated Poisson random measure are denoted by $\nu(de, dt) = \lambda(de)dt$ and $\tilde{\mu}(de, dt) = \mu(de, dt) - \lambda(de)dt$ respectively, such that $\{\tilde{\mu}(A \times [0, t]) = (\mu - \nu)(A \times [0, t])\}_{0 \leq t \leq T}$ is a martingale for all $A \in \mathcal{E}$. We also assume that $\lambda(de)$ is a σ -finite measure on (E, \mathcal{E}) satisfying

$$\int_E (1 \wedge |e|^2) \lambda(de) < +\infty. \quad (2.4)$$

Based on the stochastic basis $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{0 \leq t \leq T}, \mathbb{P})$, we introduce the backward stochas-

tic differential equation with jumps

$$\begin{cases} \mathbf{X}_t = \mathbf{X}_0 + \int_0^t \mathbf{b}(s, \mathbf{X}_s) ds + \int_0^t \boldsymbol{\sigma}(s, \mathbf{X}_s) d\mathbf{W}_s + \int_0^t \int_E \boldsymbol{\beta}(s, \mathbf{X}_{s-}, \mathbf{e}) \tilde{\mu}(d\mathbf{e}, ds) \\ \mathbf{Y}_t = \boldsymbol{\xi} + \int_t^T \mathbf{f}(s, \mathbf{X}_s, \mathbf{Y}_s, \mathbf{Z}_s, \boldsymbol{\Gamma}_s) ds - \int_t^T \mathbf{Z}_s d\mathbf{W}_s - \int_t^T \int_E \mathbf{U}_s(\mathbf{e}) \tilde{\mu}(d\mathbf{e}, ds), \end{cases} \quad (2.5)$$

where the solution is the quadruplet $(\mathbf{X}_t, \mathbf{Y}_t, \mathbf{Z}_t, \mathbf{U}_t)$ with $\mathbf{X}_t \in \mathbb{R}^d$, $\mathbf{Y}_t \in \mathbb{R}^q$, $\mathbf{Z}_t \in \mathbb{R}^{q \times m}$ and $\mathbf{U}_t \in \mathbb{R}^q$. The map $\mathbf{b} : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ is referred to as the drift coefficient, $\boldsymbol{\sigma} : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^d \times \mathbb{R}^m$ is the local diffusion coefficient, $\boldsymbol{\beta} : [0, T] \times \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ is the jump coefficient, $\mathbf{f} : [0, T] \times \mathbb{R}^m \times \mathbb{R}^d \times \mathbb{R}^{d \times m} \times \mathbb{R}^d \rightarrow \mathbb{R}^q$ is the generator of the BSDE, and the processes $\boldsymbol{\Gamma}_s : [0, T] \rightarrow \mathbb{R}^q$ is defined by $\boldsymbol{\Gamma}_s = \int_E \mathbf{U}_s(\mathbf{e}) \eta(\mathbf{e}) \lambda(d\mathbf{e})$ for a given bounded function $\eta : \mathbb{R}^d \rightarrow \mathbb{R}$. The terminal condition $\boldsymbol{\xi}$, which is an \mathcal{F}_T -measurable random vector, is assumed to be a function of \mathbf{X}_T , denoted by $\boldsymbol{\xi} = \boldsymbol{\varphi}(\mathbf{X}_T)$. Note that the integrals in (2.5) with respect to the m -dimensional Brownian motion \mathbf{W}_t and the d -dimensional compensated Poisson measure $\tilde{\mu}(d\mathbf{e}, dt)$ are Itô-type stochastic integrals. The first equation in (2.5) is the forward SDE system, which is a general Lévy process, while the second equation is a system of BSDEs driven by \mathbf{X}_t . The quadruplet $(\mathbf{X}_t, \mathbf{Y}_t, \mathbf{Z}_t, \mathbf{U}_t)$ is called an L^2 -adapted solution if it is $\{\mathcal{F}_t\}$ -adapted, square integrable process which satisfies the BSDE in (2.5). Under standard assumptions on the given functions \mathbf{b} , $\boldsymbol{\sigma}$, \mathbf{f} , $\boldsymbol{\varphi}$ and $\boldsymbol{\beta}$, the existence and uniqueness of the solution of the system in (2.5) with a nonlinear generator \mathbf{f} have been proved in [1].

Based on the extension of the nonlinear Feynman-Kac formula for BSDEs proposed in [1], the adapted solution $(\mathbf{X}_t, \mathbf{Y}_t, \mathbf{Z}_t, \mathbf{U}_t)$ of (2.5) can be associated to the unique viscosity solution $\mathbf{v}(t, \mathbf{x}) \in \mathcal{C}([0, T] \times \mathbb{R}^d)$ of the backward Kolmogorov equation, i.e.,

$$\begin{cases} \frac{\partial \mathbf{v}}{\partial t}(t, \mathbf{x}) + \tilde{\mathcal{L}}\mathbf{v}(t, \mathbf{x}) + \mathbf{f}(t, \mathbf{x}, \mathbf{v}, \boldsymbol{\sigma} \nabla \mathbf{v}, \mathcal{B}\mathbf{v}) = 0, \text{ for } (t, \mathbf{x}) \in [0, T] \times \mathbb{R}^d \\ \mathbf{v}(T, \mathbf{x}) = \boldsymbol{\varphi}(\mathbf{x}), \text{ for } \mathbf{x} \in \mathbb{R}^d. \end{cases} \quad (2.6)$$

In (2.6), $\boldsymbol{\varphi}(\mathbf{x})$ is the terminal condition at the time $t = T$, and the second-order integral-differential operator $\tilde{\mathcal{L}}$ is of the form

$$\begin{aligned} \tilde{\mathcal{L}}\mathbf{v}(t, \mathbf{x}) = & \sum_{i=1}^d b_i(t, \mathbf{x}) \frac{\partial \mathbf{v}}{\partial x^i}(t, \mathbf{x}) + \frac{1}{2} \sum_{i,j=1}^d (\boldsymbol{\sigma} \boldsymbol{\sigma}^\top)_{i,j}(t, \mathbf{x}) \frac{\partial^2 \mathbf{v}}{\partial x^i \partial x^j}(t, \mathbf{x}) \\ & + \int_E \left(\mathbf{v}(t, \mathbf{x} + \boldsymbol{\beta}(t, \mathbf{x}, \mathbf{e})) - \mathbf{v}(t, \mathbf{x}) - \sum_{i=1}^d \frac{\partial \mathbf{v}}{\partial x^i}(t, \mathbf{x}) \boldsymbol{\beta}(t, \mathbf{x}, \mathbf{e})^i \right) \lambda(d\mathbf{e}), \end{aligned} \quad (2.7)$$

with \mathcal{B} is an integral operator defined as

$$\mathcal{B}\mathbf{v}(t, \mathbf{x}) = \int_E [\mathbf{v}(t, \mathbf{x} + \boldsymbol{\beta}(t, \mathbf{x}, \mathbf{e})) - \mathbf{v}(t, \mathbf{x})] \eta(\mathbf{e}) \lambda(d\mathbf{e}).$$

The functions \mathbf{b} , $\boldsymbol{\sigma}$, \mathbf{f} , $\boldsymbol{\varphi}$, $\boldsymbol{\beta}$ in (2.7) have the same definitions as in the BSDE in (2.5). Under the condition that $\mathbf{X}_s = \mathbf{x}$ for a fixed $s \in [0, T]$, the solution $(\mathbf{Y}_t, \mathbf{Z}_t, \mathbf{U}_t)$ of the BSDE for $s \leq t \leq T$ can be represented by

$$\begin{cases} \mathbf{Y}_t = \mathbf{v}(t, \mathbf{X}_t), \\ \mathbf{Z}_t = \boldsymbol{\sigma}(t, \mathbf{X}_t) \nabla \mathbf{v}(t, \mathbf{X}_t), \\ \mathbf{U}_t = \mathbf{v}(t, \mathbf{X}_{t-} + \boldsymbol{\beta}(t, \mathbf{X}_{t-}, \mathbf{e})) - \mathbf{v}(t, \mathbf{X}_{t-}), \end{cases} \quad (2.8)$$

where $\nabla \mathbf{v}$ denotes the gradient of \mathbf{v} with respect to \mathbf{X}_t and $\mathbf{\Gamma}_t = \mathcal{B}\mathbf{v}(t, \mathbf{X}_t)$. By comparing the operator $\tilde{\mathcal{L}}$ in (2.7) and the BSDE in (2.5), we can see that the incremental $d\mathbf{X}_t$ consists of three components: the drift term $\mathbf{b}(t, \mathbf{X}_t)dt$, the Brownian diffusion $\boldsymbol{\sigma}(t, \mathbf{X}_t)d\mathbf{W}_t$, and the jump diffusion $\int_E \boldsymbol{\beta}(t, \mathbf{X}_{t-}, \mathbf{e})\tilde{\mu}(d\mathbf{e}, dt)$.

Now let us relate the BSDE in (2.5) to the nonlocal diffusion equation in (2.3). Without loss of generality, we only consider the case when \mathbf{Y}_t is a scalar function, i.e., by setting $q = 1$ in (2.5); the numerical schemes and analysis presented in this paper can be directly extended to the case of $q > 1$. Observing that the operator $\tilde{\mathcal{L}}$ in (2.7) is a generalization of the nonlocal operator \mathcal{L} in (2.3), we aim to find a simplified form of the BSDE which is the stochastic representation of the nonlocal diffusion equation in (2.3). To proceed, we define $\lambda(d\mathbf{e}) = \gamma(\mathbf{e})d\mathbf{e}$ for $\mathbf{e} \in E$ and $\lambda(d\mathbf{e}) = 0$ for $\mathbf{e} \notin E$, satisfying the condition in (2.4). Due to the integrability assumption of $\gamma(\mathbf{e})$ in (2.2), the jump diffusion term in (2.5) is a *compensated compound Poisson process*. In this case, the compensated Poisson random measure $\tilde{\mu}(d\mathbf{e}, dt)$ can be represented by

$$\tilde{\mu}(d\mathbf{e}, dt) = \mu(d\mathbf{e}, dt) - \lambda\rho(\mathbf{e})d\mathbf{e}dt, \quad (2.9)$$

where $\lambda(d\mathbf{e}) = \lambda\rho(\mathbf{e})d\mathbf{e}$, with λ being the intensity of Poisson jumps and $\rho(\mathbf{e})d\mathbf{e}$ being the probability measure of the jump size $\boldsymbol{\beta}(t, \mathbf{x}, \mathbf{e})$ satisfying $\int_E \rho(\mathbf{e})d\mathbf{e} = 1$. Then, the operator $\tilde{\mathcal{L}}$ in (2.7) can be simplified to the nonlocal operator \mathcal{L} in (2.1) by setting $m = 0$, i.e., removing the Brownian motion, and

$$\begin{aligned} \boldsymbol{\sigma} &= 0, \quad \boldsymbol{\beta} = \mathbf{e}\mathcal{I}_D(\mathbf{e}), \quad \lambda = \int_E \gamma(\mathbf{e})d\mathbf{e}, \\ \rho(\mathbf{e}) &= \frac{1}{\lambda}\gamma(\mathbf{e}), \quad b_i = \int_E \mathbf{e}\gamma(\mathbf{e})d\mathbf{e} \quad \text{for } i = 1, \dots, d, \end{aligned} \quad (2.10)$$

where $\mathcal{I}_D(\mathbf{e})$ is the characteristic function of the domain D . In the context of $\boldsymbol{\beta} = \mathbf{e}\mathcal{I}_D(\mathbf{e})$, the drift coefficient \mathbf{b} is a constant vector which indeed helps cancel the first-order derivatives of u in $\tilde{\mathcal{L}}$. On the other hand, substituting $\gamma(\mathbf{e})$ into the definition of \mathbf{b} , we have

$$\int_0^t b_i ds = \lambda \int_0^t \int_E \mathbf{e}\rho(\mathbf{e})d\mathbf{e}ds = \lambda t\mathbb{E}[\mathbf{e}] \quad \text{for } i = 1, \dots, d, \quad (2.11)$$

which is the compensator of \mathbf{X}_t . Hence, \mathbf{X}_t is just a compound Poisson process under the Poisson measure $\mu(d\mathbf{e}, dt)$ without compensation. Then, by defining $f(t, \mathbf{x}, u) = g(T - t, \mathbf{x}, u)$ and $\varphi(\mathbf{x}) = u_0(\mathbf{x})$, with g and u_0 being the forcing term and initial condition in (2.3) respectively, we obtain the BSDE corresponding to the nonlocal diffusion equation in (2.3)

$$\begin{cases} \mathbf{X}_t = \mathbf{X}_0 + \int_0^t \int_E \mathbf{e} \mu(d\mathbf{e}, ds), \\ Y_t = \varphi(\mathbf{X}_T) + \int_t^T f(s, \mathbf{X}_s, Y_s)ds - \int_t^T \int_E U_s(\mathbf{e}) \tilde{\mu}(d\mathbf{e}, ds). \end{cases} \quad (2.12)$$

According to Theorem 2.1 in [1], the well-posedness of the BSDE in (2.12) requires the generator f is globally Lipschitz continuous, i.e., there exists $K > 0$ such that

$$|f(t, \mathbf{x}, y) - f(t, \mathbf{x}', y')| \leq K(|\mathbf{x} - \mathbf{x}'| + |y - y'|)$$

for all $0 \leq t \leq T$ with $\mathbf{x}, \mathbf{x}' \in \mathbb{R}^d$ and $y, y' \in \mathbb{R}$. In this case, there exists a unique solution $(Y_t, U_t) \in S^2 \times L^2(\tilde{\mu})$ where S^2 is the set of \mathcal{F}_t -adapted càdlàg processes $\{Y_t, 0 \leq t \leq T\}$ such that

$$\|Y_t\|_{S^2}^2 = \mathbb{E} \left[\left(\sup_{0 \leq t \leq T} |Y_t| \right)^2 \right] < \infty,$$

and $L^2(\tilde{\mu})$ is the set of mappings $U : \Omega \times [0, T] \times E \rightarrow \mathbb{R}$ such that

$$\|U_t\|_{L^2(\tilde{\mu})}^2 = \mathbb{E} \left[\int_0^T \int_E U_t(e)^2 \lambda(de) dt \right] < \infty.$$

To relate the solution (Y_t, U_t) to the viscosity solution u of the nonlocal diffusion equation in (2.3), we assume φ and f are continuous functions satisfying

$$|f(t, \mathbf{x}, 0)| \leq C(1 + |\mathbf{x}|^p) \quad \text{and} \quad |\varphi(\mathbf{x})| \leq C(1 + |\mathbf{x}|^p), \quad (t, \mathbf{x}) \in [0, T] \times \mathbb{R}^d,$$

for some real $C, p > 0$. Then, under the condition that $\mathbf{X}_s = \mathbf{x}$ for a fixed $s \in [0, T]$, the solution (Y_t, U_t) for $s \leq t \leq T$ can be represented by

$$Y_t = u(T - t, \mathbf{X}_t) \quad \text{and} \quad U_t = u(T - t, \mathbf{X}_{t-} + \mathbf{e}) - u(T - t, \mathbf{X}_{t-}),$$

where $u \in \mathcal{C}([0, T] \times \mathbb{R}^d)$ is the viscosity solution of (2.3) satisfying

$$|u(t, \mathbf{x})| \leq C(1 + |\mathbf{x}|^p), \quad (t, \mathbf{x}) \in [0, T] \times \mathbb{R}^d,$$

again, for some real $C, p > 0$. Moreover, if φ and f are bounded and uniformly continuous, then u is also bounded and uniformly continuous.

Now we introduce the following notations which will be used throughout. Let $\mathcal{F}_s^{t, \mathbf{x}}$ ($t \leq s \leq T$) be a σ -field generated by the stochastic process $\{\mathbf{x} + \mathbf{X}_r - \mathbf{X}_t, t \leq r \leq s\}$ starting from the time-space point (t, \mathbf{x}) , and set $\mathcal{F}^{t, \mathbf{x}} = \mathcal{F}_T^{t, \mathbf{x}}$. Denote by $\mathbb{E}[\cdot]$ the mathematical expectation and $\mathbb{E}_s^{t, \mathbf{x}}[\cdot]$ the conditional mathematical expectation under the σ -field $\mathcal{F}_s^{t, \mathbf{x}}$ ($t \leq s \leq T$), i.e., $\mathbb{E}_s^{t, \mathbf{x}}[\cdot] = \mathbb{E}[\cdot | \mathcal{F}_s^{t, \mathbf{x}}]$. Particularly, for the sake of notational convenience, when $s = t$, we use $\mathbb{E}_t^{\mathbf{x}}[\cdot]$ to denote $\mathbb{E}[\cdot | \mathcal{F}_t^{t, \mathbf{x}}]$, i.e., the mathematical expectation under the condition that $\mathbf{X}_t = \mathbf{x}$.

3. Numerical schemes for BSDEs with jumps. In this section we propose numerical schemes for the BSDE in (2.12) in order to solve the nonlocal diffusion equation in (2.3). Specifically, a semi-discrete scheme for time discretization is studied in §3.1, and a fully-discrete scheme is constructed in §3.2 by incorporating appropriate spatial discretization techniques.

3.1. The semi-discrete scheme. For the time interval $[0, T]$, we introduce the partition

$$\mathcal{T}_\tau = \{0 = t_0 < \dots < t_N = T\} \quad (3.1)$$

with $\Delta t_n = t_{n+1} - t_n$ and $\tau = \max_{0 \leq n \leq N-1} \Delta t_n$. In the time interval $[t_n, t_{n+1}]$ for $0 \leq n \leq N-1$, under the condition that $\mathbf{X}_{t_n} = \mathbf{x} \in \mathbb{R}^d$, the BSDE in (2.12) can be rewritten as

$$\mathbf{X}_s = \mathbf{x} + \int_{t_n}^s \int_E \mathbf{e} \mu(de, dr) \quad \text{for } t_n \leq s \leq t_{n+1}, \quad (3.2)$$

$$Y_{t_n} = Y_{t_{n+1}} + \int_{t_n}^{t_{n+1}} f(s, \mathbf{X}_s, Y_s) ds - \int_{t_n}^{t_{n+1}} \int_E U_s(\mathbf{e}) \tilde{\mu}(d\mathbf{e}, ds). \quad (3.3)$$

Taking the conditional mathematical expectation $\mathbb{E}_{t_n}^{\mathbf{x}}[\cdot]$ on both sides of (3.3), due to the fact that $\int_{t_n}^t \int_E U_s(\mathbf{e}) \tilde{\mu}(d\mathbf{e}, ds)$ for $t > t_n$ is a martingale, we obtain that

$$Y_{t_n} = \mathbb{E}_{t_n}^{\mathbf{x}}[Y_{t_{n+1}}] + \int_{t_n}^{t_{n+1}} \mathbb{E}_{t_n}^{\mathbf{x}}[f(s, \mathbf{X}_s, Y_s)] ds, \quad (3.4)$$

where $\mathbb{E}_{t_n}^{\mathbf{x}}[Y_{t_n}] = Y_{t_n}(\mathbf{x}) = u(T - t_n, \mathbf{x})$, $Y_{t_{n+1}} = Y_{t_{n+1}}(\mathbf{X}_{t_{n+1}}) = u(T - t_{n+1}, \mathbf{X}_{t_{n+1}})$ and the integrand $\mathbb{E}_{t_n}^{\mathbf{x}}[f(s, \mathbf{X}_s, Y_s)]$ is a deterministic function of $s \in [t_n, t_{n+1}]$. To estimate the integral in (3.4), several numerical methods for approximating integrals with deterministic integrands can be used. Here, we utilize the θ -scheme proposed in [20, 21], which yields

$$Y_{t_n} = \mathbb{E}_{t_n}^{\mathbf{x}}[Y_{t_{n+1}}] + (1 - \theta)\Delta t_n \mathbb{E}_{t_n}^{\mathbf{x}}[f(t_{n+1}, \mathbf{X}_{t_{n+1}}, Y_{t_{n+1}})] \\ + \theta\Delta t_n f(t_n, \mathbf{X}_{t_n}, Y_{t_n}) + R_n, \quad (3.5)$$

where $\theta \in [0, 1]$ and the residual R_n is given by

$$R_n = \int_{t_n}^{t_{n+1}} \left\{ \mathbb{E}_{t_n}^{\mathbf{x}}[f(s, \mathbf{X}_s, Y_s)] - \theta f(t_n, \mathbf{X}_{t_n}, Y_{t_n}) \right. \\ \left. - (1 - \theta)\mathbb{E}_{t_n}^{\mathbf{x}}[f(t_{n+1}, \mathbf{X}_{t_{n+1}}, Y_{t_{n+1}})] \right\} ds. \quad (3.6)$$

Next we propose a semi-discrete scheme for the BSDE in (2.12) as follows: given the random variable $Y_N = Y_{t_N} = \varphi(\mathbf{X}_T)$ as the terminal condition, for $n = N - 1, \dots, 1, 0$, under the condition that $\mathbf{X}_n = \mathbf{x}$, the solution Y_{t_n} is approximated by Y_n satisfying

$$\begin{cases} \mathbf{X}_{n+1} = \mathbf{X}_n + \int_E \mathbf{e} \mu(d\mathbf{e}, \Delta t), \\ Y_n = \mathbb{E}_{t_n}^{\mathbf{x}}[Y_{n+1}] + (1 - \theta)\Delta t_n \mathbb{E}_{t_n}^{\mathbf{x}}[f(t_{n+1}, \mathbf{X}_{n+1}, Y_{n+1})] \\ \quad + \theta\Delta t_n f(t_n, \mathbf{X}_n, Y_n), \end{cases} \quad (3.7)$$

where $0 \leq \theta \leq 1$. By choosing different values for θ , we obtain different semi-discrete schemes, e.g., $\theta = 0$, $\theta = 1$ and $\theta = \frac{1}{2}$ lead to forward Euler (FE), backward Euler (BE) and Crank-Nicolson (CN) schemes, respectively.

Note that since \mathbf{X}_t is the standard compound Poisson process, the probability distributions of \mathbf{X}_t or any incremental $\mathbf{X}_t - \mathbf{X}_{t'}$, with $0 \leq t' < t \leq T$, are well known. Thus, one does not need to discretize \mathbf{X}_t so that, in (3.7), \mathbf{X}_{n+1} denotes the exact solution evaluated at t_{n+1} . In this case, R_n in (3.6) is the local truncation error of the scheme (3.7); a rigorous error analysis of R_n is given in §4. Other time-stepping schemes, e.g., linear multi-step schemes [22], can be directly used in order to further improve the accuracy of time discretization. Note that in the general case where the coefficients \mathbf{b} and $\boldsymbol{\sigma}$ in (2.5) and (2.7) are functions of t and \mathbf{X}_t , another time-stepping scheme [18] is also needed for to discretize \mathbf{X}_t , so that an extra local truncation error will be introduced into the semi-discrete scheme (3.7). This is beyond the scope of this effort but will be the focus of future work.

3.2. The fully-discrete scheme. The semi-discrete scheme (3.7) is not computationally feasible, because the involved conditional mathematical expectation $\mathbb{E}_{t_n}^{\mathbf{x}}[\cdot]$ is defined over the whole real space \mathbb{R}^d . Thus, an effective spatial discretization approach is also necessary in order to approximate $\mathbb{E}_{t_n}^{\mathbf{x}}[\cdot]$. To proceed, we partition the d -dimensional Euclidean space \mathbb{R}^d by

$$\mathcal{S}_h^d = \mathcal{S}_{h_1}^1 \times \mathcal{S}_{h_2}^2 \times \cdots \times \mathcal{S}_{h_d}^d,$$

where $\mathcal{S}_{h_k}^k$ for $k = 1, \dots, d$ is a partition of the one-dimensional real space \mathbb{R} , i.e.,

$$\mathcal{S}_{h_k}^k = \left\{ x_i^k \mid x_i^k \in \mathbb{R}, i \in \mathbb{Z}, x_i^k < x_{i+1}^k, \lim_{i \rightarrow +\infty} x_i^k = +\infty, \lim_{i \rightarrow -\infty} x_i^k = -\infty \right\},$$

such that $h_k = \max_{i \in \mathbb{Z}} \{ |x_i^k - x_{i-1}^k| \}$ and $h = \max_{1 \leq k \leq d} h_k$. For each multi-index $\mathbf{i} = (i_1, i_2, \dots, i_d) \in \mathbb{Z}^d$, the corresponding grid point in \mathcal{S}_h^d is denoted by $\mathbf{x}_i = (x_{i_1}^1, \dots, x_{i_d}^d)$.

Now we turn to constructing a quadrature rule for approximating $\mathbb{E}_{t_n}^{\mathbf{x}_i}[\cdot]$ in (3.7) for $(t_n, \mathbf{x}_i) \in \mathcal{T}_\tau \times \mathcal{S}_h^d$. Such expectation is defined with respect to the probability distribution of the incremental stochastic process $\Delta \mathbf{X}_{n+1} = \mathbf{X}_{n+1} - \mathbf{X}_n = \mathbf{X}_{n+1} - \mathbf{x}_i$, which is a compound Poisson process starting from the grid point (t_n, \mathbf{x}_i) . Here we take $\mathbb{E}_{t_n}^{\mathbf{x}_i}[Y_{n+1}]$ as an example and the proposed quadrature rule can be directly extended to approximate $\mathbb{E}_{t_n}^{\mathbf{x}_i}[f(t_{n+1}, \mathbf{X}_{n+1}, Y_{n+1})]$. It is well known that the number of jumps of \mathbf{X}_t within $(t_n, t_{n+1}]$, denoted by N_t , follows a Poisson distribution with intensity $0 < \lambda < +\infty$, and the size of each Poisson jump follows the distribution $\rho(e)de$ defined in (2.10). Thus, $\mathbb{E}_{t_n}^{\mathbf{x}_i}[Y_{n+1}]$ can be represented by

$$\begin{aligned} \mathbb{E}_{t_n}^{\mathbf{x}_i}[Y_{n+1}] &= \sum_{m=0}^{\infty} \mathbb{P}\{N_{t_{n+1}} - N_{t_n} = m\} \mathbb{E}\left[Y_{n+1}\left(\mathbf{x}_i + \sum_{k=1}^m \mathbf{e}_k\right)\right] \\ &= \sum_{m=0}^{\infty} \exp(-\lambda \Delta t_n) \frac{(\lambda \Delta t_n)^m}{m!} \mathbb{E}\left[Y_{n+1}\left(\mathbf{x}_i + \sum_{k=1}^m \mathbf{e}_k\right)\right] \\ &= \exp(-\lambda \Delta t_n) Y_{n+1}(\mathbf{x}_i) + \sum_{m=1}^{\infty} \exp(-\lambda \Delta t_n) \frac{(\lambda \Delta t_n)^m}{m!} \\ &\quad \times \int_E \cdots \int_E Y_{n+1}\left(\mathbf{x}_i + \sum_{k=1}^m \mathbf{e}_k\right) \left(\prod_{k=1}^m \rho(\mathbf{e}_k)\right) d\mathbf{e}_1 \cdots d\mathbf{e}_m, \end{aligned} \tag{3.8}$$

where $\mathbf{e}_k = (e_k^1, \dots, e_k^d)$ for $k = 1, \dots, m$ is the size of the k -th jump. Observing that the probability of having m jumps within $(t_n, t_{n+1}]$ is of order $\mathcal{O}((\lambda \Delta t_n)^m)$, the conditional mathematical expectation $\mathbb{E}_{t_n}^{\mathbf{x}_i}[Y_{n+1}]$ can be approximated by a truncation of (3.8), i.e., the sum of a *finite* sequence, where the number of terms retained is determined according to the prescribed accuracy. For example, if we take $\theta = \frac{1}{2}$ in (3.7), we expect a second-order convergence from the time discretization, i.e., the local truncation error R_n in (3.5) should be of order $\mathcal{O}((\Delta t_n)^3)$. In this case, assuming the intensity value λ is of order $\mathcal{O}(1)$, the first three terms should be retained in (3.8) in order to guarantee the error from the truncation of (3.8) matches the order of R_n . Hence, in the sequel, we denote by $\mathbb{E}_{t_n, M_y}^{\mathbf{x}_i}[Y_{n+1}]$ the approximation of $\mathbb{E}_{t_n}^{\mathbf{x}_i}[Y_{n+1}]$ by retaining the first $M_y + 1$ terms, where M_y indicates the number of jumps included in $\mathbb{E}_{t_n, M_y}^{\mathbf{x}_i}[Y_{n+1}]$. An analogous notation $\mathbb{E}_{t_n, M_f}^{\mathbf{x}_i}[f_{n+1}]$ is used to represent the approximation of $\mathbb{E}_{t_n}^{\mathbf{x}_i}[f_{n+1}]$ by retaining M_f jumps.

Next, we also need to approximate the $m \times d$ -dimensional integral in (3.8) for $m = 1, \dots, M_y$. This can be accomplished by selecting an appropriate quadrature rule base on the properties of $\rho(\mathbf{e})$ and the smoothness of Y_{n+1} . A straightforward choice is to utilize Monte Carlo methods by drawing samples from $\prod_{k=1}^m \rho(\mathbf{e}_k)$, but this is inefficient because of slow convergence. To enhance the convergence rate, an alternative way is to use the tensor product of high-order one-dimensional quadrature rules, e.g., Newton-Cotes, Gaussian, etc.. In addition, for multi-dimensional problems, i.e., $d > 1$, by requiring second-order convergence $\mathcal{O}((\Delta t)^2)$, i.e., $M_y = 2$, sparse-grid quadrature rules [4, 12] can be applied to further improve the computational efficiency. Without loss of generality, for $m = 1, \dots, M_y$, we denote by $\{w_q^m\}_{q=1}^{Q_m}$ and $\{\mathbf{a}_q^{1,m}, \dots, \mathbf{a}_q^{m,m}\}_{q=1}^{Q_m}$ the set of weights and points, respectively, of the selected quadrature rule for estimating the m -th integral in (3.8). Then, the approximation of $\mathbb{E}_{t_n}^{\mathbf{x}_i}[Y_{n+1}]$, denoted by $\widehat{\mathbb{E}}_{t_n, M_y}^{\mathbf{x}_i}[Y_{n+1}]$ is given by

$$\begin{aligned} \widehat{\mathbb{E}}_{t_n, M_y}^{\mathbf{x}_i}[Y_{n+1}] &= \exp(-\lambda \Delta t) Y_{n+1}(\mathbf{x}_i) \\ &+ \sum_{m=1}^{M_y} \exp(-\lambda \Delta t) \frac{(\lambda \Delta t)^m}{m!} \sum_{q=1}^{Q_m} w_q^m Y_{n+1}\left(\mathbf{x}_i + \sum_{k=1}^m \mathbf{a}_q^{k,m}\right). \end{aligned} \quad (3.9)$$

Note that it is possible that the quadrature points $\{\mathbf{x}_i + \sum_{k=1}^m \mathbf{a}_q^{k,m}\}_{q=1}^{Q_m}$ in (3.9) do not belong to the spatial grid \mathcal{S}_h^d . The simplest approach to manage this difficult is to add all quadrature points to the spatial grid at each time stage. However, this will result in an exponential growth of the total number of grid points as the number of time steps increases. Thus, we follow the same strategy as in [20, 21, 22], and construct piecewise Lagrange interpolating polynomials based on \mathcal{S}_h to approximate Y_{n+1} at the non-grid points. Specifically, at any given point $\mathbf{x} = (x^1, \dots, x^d) \in \mathbb{R}^d$, $Y_{n+1}(\mathbf{x})$ is approximated by

$$Y_{n+1}(\mathbf{x}) \approx Y_{n+1,p}(\mathbf{x}) = \sum_{j_1=1}^{p+1} \cdots \sum_{j_d=1}^{p+1} \left(Y_{n+1,p}^{i_{j_1}, \dots, i_{j_d}} \prod_{k=1}^d \prod_{\substack{1 \leq j \leq p+1 \\ j \neq j_k}} \frac{x^k - x_{i_j}^k}{x_{i_{j_k}}^k - x_{i_j}^k} \right),$$

where $Y_{n+1,p}(\mathbf{x})$ is a p -th order tensor-product Lagrange interpolating polynomial. For $k = 1, \dots, d$, the interpolation points $\{x_{i_j}^k\}_{j=1}^{p+1} \subset \mathcal{S}_h^k$ are the closest $p+1$ neighboring points of x^k such that $(x_{i_{j_1}}^1, \dots, x_{i_{j_d}}^d)$, for $j_k = 1, \dots, p+1$ and $k = 1, \dots, d$, constitute a local tensor-product sub-grid around \mathbf{x} . The fully-discrete solution at the interpolation point $(x_{i_{j_1}}^1, \dots, x_{i_{j_d}}^d) \in \mathcal{S}_h^d$ is given by $Y_{n+1,p}^{i_{j_1}, \dots, i_{j_d}}$. Note that $Y_{n+1,p}(\mathbf{x})$ does not interpolate the semi-discrete solution $Y_{n+1}(\mathbf{x})$ due to the error $Y_{n+1}(\mathbf{x}_i) - Y_{n+1,p}^{\mathbf{i}}$. Therefore, the fully-discrete scheme of the BSDE in (2.12) is described as follows: given the random variable $Y_N(\mathbf{x}_i)$ for $\mathbf{i} \in \mathbb{Z}^d$, and for $n = N-1, \dots, 0$, solve the quantities $Y_{n,p}^{\mathbf{i}}$ for $\mathbf{i} \in \mathbb{Z}^d$ such that

$$\begin{cases} \mathbf{X}_{n+1} = \mathbf{x}_i + \int_E \mathbf{e} \mu(d\mathbf{e}, \Delta t), \\ Y_{n,p}^{\mathbf{i}} = \widehat{\mathbb{E}}_{t_n, M_y}^{\mathbf{x}_i}[Y_{n+1,p}] + (1-\theta)\Delta t_n \widehat{\mathbb{E}}_{t_n, M_f}^{\mathbf{x}_i}[f(t_{n+1}, \mathbf{X}_{n+1}, Y_{n+1,p})] \\ \quad + \theta \Delta t_n f(t_n, \mathbf{X}_n, Y_{n,p}^{\mathbf{i}}), \end{cases} \quad (3.10)$$

where $\theta \in [0, 1]$ and the non-negative integers M_y, M_f indicate the Poisson jumps included in the approximations of $\mathbb{E}_{t_n}^{\mathbf{x}_i}[Y_{n+1,p}]$ and $\mathbb{E}_{t_n}^{\mathbf{x}_i}[f]$, respectively.

Note that, by using the quadrature rule, the approximation in (3.9) is defined in a bounded domain for a fixed $\mathbf{x}_i \in \mathcal{S}_h^d$. As such, the approximate solution $Y_{0,p}$ in any bounded subdomain of \mathbb{R}^d can be obtained by solving the fully-discrete scheme (3.10) in a bounded subset of \mathcal{S}_h^d . We also observe from (3.10) that at each point $(t_n, x_i) \in \mathcal{T}_\tau \times \mathcal{S}_h^d$, the computation of $Y_{n,p}^i$ only depends on \mathbf{X}_{n+1} and $Y_{n+1,p}$ even though an implicit time-stepping scheme ($\theta > 0$) is used. This means $\{Y_{n,p}^i\}_{i \in \mathbb{Z}}$ on each time stage can be computed independently, so that the difficulty of solving linear systems with possibly dense matrices is completely avoided. Instead, $Y_{n,p}^i$ can be either computed explicitly (in case f is linearly dependent on $Y_{n,p}^i$), or obtained by solving a nonlinear equation (in case f is nonlinear with respect to $Y_{n,p}^i$). This feature makes it straightforward to develop massively parallel algorithms, which are, in terms of scalability, very similar to the CTRW method. Moreover, the scheme (3.10) is expected to outperform the CRTW method because it achieves high-order accuracy (discussed in §4) from the discretization and is also capable of solving problems that include a general nonlinear forcing term g . In fact, the combination of accurate approximations and scalable computations are the key advantages of our approach, compared to the existing methods for the nonlocal diffusion problem in (2.3).

REMARK 1. *The nonlinear Feynman-Kac formula converted the effect of the nonlocal operator \mathcal{L} in (2.1) to the behavior of the corresponding compound Poisson process, which is accurately approximated by the conditional mathematical expectations $\mathbb{E}_{t_n}^{\mathbf{x}}[\cdot]$. As such, there is no need to discretize the nonlocal operator in the fully-discrete scheme (3.10), so that stability does not require any CFL-type restriction on the time step; this is in contrast to classic numerical schemes for which explicit schemes do require a CFL condition. Moreover, according to the analysis in [20, 22], the approximation (3.10) is stable as long as the semi-discrete scheme is absolutely stable.*

REMARK 2. *The total computational cost of the scheme (3.10) is dominated by the cost of approximating $\mathbb{E}_{t_n}^{\mathbf{x}_i}[\cdot]$ at each grid point $\mathbf{x}_i \in \mathcal{S}_h^d$, using the formula in (3.9). For example, when solving a three-dimensional problem ($d = 3$) and retaining two Lévy jumps ($M_y = M_f = 2$), our approach requires the approximation of multiple six-dimensional integrals. In this case, sparse-grid quadrature rules [19] can be used to alleviate the explosion in computational cost coming from the high-dimensional systems.*

4. Error estimates. In this section, we analyze both the semi-discrete and fully-discrete schemes, given by (3.7) and (3.10) respectively, for the BSDE in (2.12). Without loss of generality, we only consider the one-dimensional case ($d = 1$) and set $\theta = \frac{1}{2}$; the following analysis can be directly extended to multi-dimensional cases. For simplicity, we also assume \mathcal{T}_τ and \mathcal{S}_h^d are both uniform grids, i.e., $\Delta t = \Delta t_n$ for $n = 1, \dots, N$ and $\Delta x = x_i - x_{i-1}$ for $i \in \mathbb{Z}$.

Note that the well-posedness of the BSDE in (2.12) only requires global Lipschitz continuity on f with respect to X_t and Y_t . However, in order to obtain error estimates, we need to impose stronger regularity on f . To this end, we first introduce the

following notation:

$$\begin{aligned} & \mathcal{C}_b^k(D_1 \times \cdots \times D_J) \\ &= \left\{ \psi : \prod_{j=1}^J D_j \rightarrow \mathbb{R} \mid \frac{\partial^{|\alpha|} \psi}{\partial \mathbf{x}^\alpha} \text{ is bounded and continuous for } 0 \leq |\alpha| \leq k \right. \\ & \quad \left. \text{where } \alpha = (\alpha_1, \dots, \alpha_J) \text{ with } \{\alpha_j\}_{j=1}^J \subset \mathbb{N} \text{ and } |\alpha| = \alpha_1 + \cdots + \alpha_J \right\}, \end{aligned}$$

where $D_1 \times \cdots \times D_J \subset \mathbb{R}^J$. Next we present the following lemma that relates the smoothness of $f(t, x, y)$ to the regularity of the solution $u(t, x)$ of the nonlocal diffusion equation in (2.3).

LEMMA 1. *Under assumptions in (2.2) and (2.4), if $\varphi(x) \in \mathcal{C}_b^k(\mathbb{R})$, $f(t, x, y) \in \mathcal{C}_b^k([0, T] \times \mathbb{R}^2)$ in (2.12) and $\frac{\partial^k f}{\partial x^{\alpha_1} \partial y^{\alpha_2}}$ with $\alpha_1 + \alpha_2 \leq k$ is uniformly Lipschitz continuous with respect to x and y , then the nonlocal diffusion equation in (2.3) with $d = 1$, $u_0(x) = \varphi'(x)$ and $g(t, x, u) = f(T - t, x, u)$ has a unique viscosity solution $u(t, x) \in \mathcal{C}_b^k([0, T] \times \mathbb{R})$.*

Proof. Here we only prove the case of $k = 1$, as the cases $k > 1$ can be proved by recursively using the following argument. Based on the regularity of f , it is easy to see that $\frac{\partial u}{\partial t}$ is bounded and continuous with respect to t , so that we focus on proving the regularity of $\frac{\partial u}{\partial x}$. Let ∇X_t , ∇Y_t and ∇U_t be the variations of the stochastic processes X_t , Y_t and U_t with respect to X_t in (2.12), respectively. Then the triple $(\nabla X_t, \nabla Y_t, \nabla U_t)$ is the solution of the following BSDE.

$$\begin{cases} \nabla X_t = 1, \\ \nabla Y_t = \varphi'(X_T) + \int_t^T f'_x(s, X_s, Y_s) + f'_y(s, X_s, Y_s) \nabla Y_s ds \\ \quad - \int_t^T \int_E \nabla U_s(e) \tilde{\mu}(de, ds), \end{cases} \quad (4.1)$$

which is a linear equation with respect to ∇X_s , ∇Y_s and ∇U_s . Due to the regularity of f , the BSDE in (4.1) has a unique solution $(\nabla Y_t, \nabla U_t)$ where ∇Y_t is bounded and continuous. By the relationship between BSDEs and PIDEs, we see that $\nabla Y_t = w(t, X_t)$ where $w(t, x)$ is the unique viscosity solution of the following PIDE [1]

$$\begin{cases} \frac{\partial w}{\partial t}(t, x) + \mathcal{L}w(t, x) + f'_x(t, x, v) + f'_v(t, x, v)w = 0 & (t, x) \in [0, T] \times \mathbb{R}, \\ w(T, x) = \varphi'(x), & \text{for } x \in \mathbb{R}, \end{cases} \quad (4.2)$$

where $v = u(T - t, x)$ is the transformed solution of the nonlocal problem in (2.3) with $d = 1$, $u_0 = \varphi'(x)$ and $g(t, x, u) = f(T - t, x, u)$. Then, by differentiating the equations in (2.3) with respect to x and comparing the resulting equations with (4.2), we can see that $w(t, x) = \frac{\partial u}{\partial x}(T - t, x)$. Due to the continuity and boundedness of $w(t, x)$, we have that $u(t, x) \in \mathcal{C}_b^1([0, T] \times \mathbb{R})$, which completes the proof. \square

Next we provide the estimate of the local truncation error R_n in (3.6).

THEOREM 1. *Using Lemma 1 with $k = 2$, the local truncation error R_n in (3.5), for the semi-discrete scheme (3.7) with $\theta = \frac{1}{2}$ can be bounded by*

$$\mathbb{E}[|R_n|] \leq C(\Delta t)^3 \text{ for } n = 0, \dots, N - 1, \quad (4.3)$$

where the constant C depends only on the terminal time T , the jump intensity λ , the upper bounds of f , φ , and their derivatives.

Proof. The formula (2.8) shows that the solution Y_t of the BSDE (2.12) can be represented by $Y_t = u(t, X_t)$. Under Lemma 1, if $f \in \mathcal{C}_b^2([0, T] \times \mathbb{R}^2)$ and $\varphi \in \mathcal{C}_b^2(\mathbb{R})$, then $u(t, X_t) \in \mathcal{C}_b^2([0, T] \times \mathbb{R})$. Thus, by defining $F(t, X_t) = f(t, X_t, u(T - t, X_t))$, we have $F \in \mathcal{C}_b^2([0, T] \times \mathbb{R})$. Before estimating the residual R_n in (3.5), we define the differential operators L^0 and L^1 by

$$\begin{cases} L^0 F(t, X_t) = \frac{\partial F}{\partial t}(t, X_t) + b \frac{\partial F}{\partial x}(t, X_t) \\ \quad + \int_E \left[F(t, X_{t-} + e) - F(t, X_{t-}) - \frac{\partial F}{\partial x}(t, X_{t-})e \right] \lambda(de), \\ L^1 F(t, X_t) = F(t, X_{t-} + e) - F(t, X_{t-}). \end{cases} \quad (4.4)$$

Based on the definition of X_s in (2.12), the integral form of the Itô formula of $F(s, X_s)$ for $t_n \leq s \leq t_{n+1}$, under the condition $X_{t_n} = x$, is given by

$$F(s, X_s) = F(t_n, x) + \int_{t_n}^s L^0 F(r, X_r) dr + \int_{t_n}^s \int_E L^1 F(r, X_{r-}) \tilde{\mu}(de, dr). \quad (4.5)$$

Thus, substituting the above formula into $\int_{t_n}^{t_{n+1}} \mathbb{E}_{t_n}^x [F(s, X_s)] ds$, we obtain

$$\begin{aligned} & \int_{t_n}^{t_{n+1}} \mathbb{E}_{t_n}^x [F(s, X_s)] ds \\ &= \int_{t_n}^{t_{n+1}} \mathbb{E}_{t_n}^x \left[F(t_n, x) + \int_{t_n}^s L^0 F(r, X_r) dr + \int_{t_n}^s L^1 F(r, X_{r-}) \tilde{\mu}(de, dr) \right] ds \\ &= F(t_n, x) \int_{t_n}^{t_{n+1}} ds + \int_{t_n}^{t_{n+1}} \int_{t_n}^s \mathbb{E}_{t_n}^x [L^0 F(r, X_r)] dr ds \\ &= F(t_n, x) \Delta t + \int_{t_n}^{t_{n+1}} \int_{t_n}^s \mathbb{E}_{t_n}^x \left[L^0 F(t_n, x) + \int_{t_n}^r L^0 L^0 F(z, X_z) dz \right] dr ds \\ &= F(t_n, x) \Delta t + \frac{1}{2} (\Delta t)^2 L^0 F(t_n, x) + \int_{t_n}^{t_{n+1}} \int_{t_n}^s \int_{t_n}^r \mathbb{E}_{t_n}^x [L^0 L^0 F(z, X_z)] dz dr ds. \end{aligned} \quad (4.6)$$

Similar derivation can be conducted to $F(t_{n+1}, X_{t_{n+1}})$ to obtain

$$\begin{aligned} & \int_{t_n}^{t_{n+1}} \mathbb{E}_{t_n}^x [F(t_{n+1}, X_{t_{n+1}})] ds \\ &= F(t_n, x) \Delta t + (\Delta t)^2 L^0 F(t_n, x) + \int_{t_n}^{t_{n+1}} \int_{t_n}^{t_{n+1}} \int_{t_n}^r \mathbb{E}_{t_n}^x [L^0 L^0 F(z, X_z)] dz dr ds. \end{aligned} \quad (4.7)$$

Therefore, the residual R_n in (3.5) with $\theta = \frac{1}{2}$ can be represented by

$$\begin{aligned} R_n &= \int_{t_n}^{t_{n+1}} \mathbb{E}_{t_n}^x [F(s, X_s)] - \frac{1}{2} F(t_n, x) - \frac{1}{2} \mathbb{E}_{t_n}^x [F(t_{n+1}, X_{t_{n+1}})] ds \\ &= \int_{t_n}^{t_{n+1}} \int_{t_n}^s \int_{t_n}^r \mathbb{E}_{t_n}^x [L^0 L^0 F(z, X_z)] dz dr ds \\ &\quad - \frac{1}{2} \int_{t_n}^{t_{n+1}} \int_{t_n}^{t_{n+1}} \int_{t_n}^r \mathbb{E}_{t_n}^x [L^0 L^0 F(z, X_z)] dz dr ds. \end{aligned} \quad (4.8)$$

Then, taking the mathematical expectation of R_n , and using Cauchy-Schwarz inequality, we have that

$$\begin{aligned}
\mathbb{E}[|R_n|] &\leq \mathbb{E} \left[\left| \int_{t_n}^{t_{n+1}} \int_{t_n}^s \int_{t_n}^r \mathbb{E}_{t_n}^x [L^0 L^0 F(z, X_z)] dz dr ds \right| \right] \\
&\quad + \frac{1}{2} \mathbb{E} \left[\left| \int_{t_n}^{t_{n+1}} \int_{t_n}^{t_{n+1}} \int_{t_n}^r \mathbb{E}_{t_n}^x [L^0 L^0 F(z, X_z)] dz dr ds \right| \right] \\
&\leq (\Delta t)^{\frac{3}{2}} \sqrt{\int_{t_n}^{t_{n+1}} \int_{t_n}^s \int_{t_n}^r \mathbb{E}[|L^0 L^0 F(z, X_z)|^2] dz dr ds} \\
&\quad + \frac{1}{2} (\Delta t)^{\frac{3}{2}} \sqrt{\int_{t_n}^{t_{n+1}} \int_{t_n}^{t_{n+1}} \int_{t_n}^r \mathbb{E}[|L^0 L^0 F(z, X_z)|^2] dz dr ds} \\
&\leq \frac{3}{2} \sup_{0 \leq t \leq T} \sqrt{\mathbb{E}[|L^0 L^0 F(t, X_t)|^2]} (\Delta t)^3 \\
&\leq C(\Delta t)^3,
\end{aligned} \tag{4.9}$$

where, by the definition of L^0 , it is easy to see that the constant C only depends on the terminal time T , the jump intensity λ the upper bounds of f , φ and their derivatives. \square

We now turn our attention to proving an error estimate for semi-discrete scheme (3.7).

THEOREM 2. *Let Y_{t_n} and Y_n , for $n = 0, 1, \dots, N$, be the exact solution of the BSDE (2.12) and the semi-discrete solution obtained by the scheme (3.7), respectively. Then, under the same conditions as that for Theorem 1, for sufficiently small time step Δt , the global truncation error $e_n = Y_{t_n} - Y_n$, for $n = N - 1, \dots, 0$, can be bounded by*

$$\mathbb{E}[|Y_{t_n} - Y_n|] \leq C(\Delta t)^2, \tag{4.10}$$

where the constant C is the same as in Theorem 1.

Proof. Given $0 \leq n \leq N - 1$ and $x \in \mathbb{R}$, subtracting (3.7) from (3.5), we have

$$\begin{aligned}
e_n &= \mathbb{E}_{t_n}^x [e_{n+1}] + \frac{\Delta t}{2} \mathbb{E}_{t_n}^x [f(t_{n+1}, X_{t_{n+1}}, Y_{t_{n+1}}) - f(t_{n+1}, X_{n+1}, Y_{n+1})] \\
&\quad + \frac{\Delta t}{2} [f(t_n, X_{t_n}, Y_{t_n}) - f(t_n, X_n, Y_n)] + R_n
\end{aligned} \tag{4.11}$$

Due to the Lipschitz continuity of f and the fact that $X_{t_{n+1}} = X_{n+1}$, $X_{t_n} = X_n = x$, we get the bound

$$|e_n| \leq \mathbb{E}_{t_n}^x [|e_{n+1}|] + \frac{L\Delta t}{2} \mathbb{E}_{t_n}^x [|e_{n+1}|] + \frac{L\Delta t}{2} |e_n| + |R_n|, \tag{4.12}$$

where L is the Lipschitz constant of f and the time step size Δt satisfies $1 - \frac{L\Delta t}{2} > 0$. Substituting the upper bound of R_n , the above inequality can be rewritten as

$$|e_n| \leq \frac{1 + \frac{L\Delta t}{2}}{1 - \frac{L\Delta t}{2}} \mathbb{E}_{t_n}^x [|e_{n+1}|] + \frac{C(\Delta t)^3}{1 - \frac{L\Delta t}{2}}. \tag{4.13}$$

Taking mathematical expectation on both sides, we obtain, by induction, an upper bound of $\mathbb{E}[|e_n|]$, i.e.,

$$\begin{aligned}\mathbb{E}[|e_n|] &\leq (1 + L\Delta t)^{N-n} \mathbb{E}[|e_N|] + C(\Delta t)^3 \frac{(1 + L\Delta t)^{N-n} - 1}{L\Delta t} \\ &\leq \frac{C(\Delta t)^2}{L} (e^{LT} - 1),\end{aligned}\quad (4.14)$$

where the constant C is the same constant from (4.9). \square

Next, we focus on analyzing the numerical error for the fully-discrete scheme (3.10). For $0 \leq n \leq N - 1$, $Y_{n,p}$ is constructed using piecewise p -th order Lagrange polynomial interpolation on the mesh \mathcal{S}_h . As for the quadrature rules involved in $\widehat{\mathbb{E}}_{t_n, M_y}^{x_i}[\cdot]$, without loss of generality, we consider a special case stated in the following assumption.

ASSUMPTION 1. For $d = 1$ and $m = 1, \dots, M_y$, the quadrature rules $\{w_q^m\}_{q=1}^{Q_m}$, $\{a_q^{1,m}, \dots, a_q^{m,m}\}_{q=1}^{Q_m}$ used in (3.9) are assumed to be the tensor-product of the selected one-dimensional quadrature rule, denoted by $\{\tilde{w}_j, \tilde{a}_j\}_{j=1}^Q$. The error in the quadrature rule is represented by

$$\begin{aligned}&\left| \int_E \cdots \int_E Y_{t_{n+1}} \left(x_i + \sum_{k=1}^m e_k \right) \left(\prod_{k=1}^m \rho(e_k) \right) de_1 \cdots de_m \right. \\ &\quad \left. - \sum_{q_1=1}^Q \cdots \sum_{q_m=1}^Q \left(\prod_{k=1}^m \tilde{w}_{q_k} \right) \cdot Y_{t_{n+1}} \left(x_i + \sum_{k=1}^m \tilde{a}_{q_k} \right) \right| \leq CQ^{-r},\end{aligned}\quad (4.15)$$

where the convergence rate $r > 0$ and the constant C are determined by the smoothness of Y_{n+1} and ρ . The same strategy is also applied to $\widehat{\mathbb{E}}_{t_n, M_f}^{x_i}[f]$.

Then, with Assumption 1, the error estimate for the fully-discrete approximation (3.10) is given in the following theorem.

THEOREM 3. Let Y_{t_n} and $Y_{n,p}^i$, for $n = 0, 1, \dots, N$, $i \in \mathbb{Z}$, be the exact solution of the BSDE (2.12) and the fully-discrete approximation obtained by the scheme (3.10) with $\theta = \frac{1}{2}$, respectively. Under Assumption 1 and the conditions in Theorems 1 and 2, the error $e_n^i = Y_{t_n}^{x_i} - Y_{n,p}^i$ can be bounded by

$$\max_{i \in \mathbb{Z}} |e_n^i| \leq C [(\Delta t)^2 + (\lambda \Delta t)^{M_y} + (\lambda \Delta t)^{M_f+1} + Q^{-r} + (\Delta x)^{p+1}], \quad (4.16)$$

where the constant C is the same constant from (4.9).

Proof. At each grid point $x_i \in \mathcal{S}_h$, substituting $X_{t_n} = x_i$ into (3.5), we have

$$Y_{t_n}^{x_i} = \mathbb{E}_{t_n}^{x_i}[Y_{t_{n+1}}] + \frac{\Delta t}{2} \mathbb{E}_{t_n}^{x_i}[f(t_{n+1}, X_{t_{n+1}}, Y_{t_{n+1}})] + \frac{\Delta t}{2} f(t_n, x_i, Y_{t_n}^{x_i}) + R_n, \quad (4.17)$$

where $Y_{t_n}^{x_i}$ is the value of the exact solution at the grid point (t_n, x_i) . Subtracting (4.17) from (3.10) leads to

$$\begin{aligned}e_n^i &= \mathbb{E}_{t_n}^{x_i}[Y_{t_{n+1}}] - \widehat{\mathbb{E}}_{t_n, M_y}^{x_i}[Y_{n+1,p}] + \frac{\Delta t}{2} \left\{ \mathbb{E}_{t_n}^{x_i}[f(t_{n+1}, X_{t_{n+1}}, Y_{t_{n+1}})] \right. \\ &\quad \left. - \widehat{\mathbb{E}}_{t_n, M_f}^{x_i}[f(t_{n+1}, X_{n+1}, Y_{n+1,p})] \right\} + \frac{\Delta t}{2} [f(t_n, x_i, Y_{t_n}^{x_i}) - f(t_n, x_i, Y_{n,p}^i)] + R_n \\ &= I_1 + I_2 + I_3 + R_n,\end{aligned}\quad (4.18)$$

where

$$\begin{aligned} I_1 &= \mathbb{E}_{t_n}^{x_i}[Y_{t_{n+1}}] - \widehat{\mathbb{E}}_{t_n, M_y}^{x_i}[Y_{n+1, p}], \\ I_2 &= \frac{\Delta t}{2} \left\{ \mathbb{E}_{t_n}^{x_i}[f(t_{n+1}, X_{t_{n+1}}, Y_{t_{n+1}})] - \widehat{\mathbb{E}}_{t_n, M_f}^{x_i}[f(t_{n+1}, X_{n+1}, Y_{n+1, p})] \right\}, \\ I_3 &= \frac{\Delta t}{2} [f(t_n, x_i, Y_{t_n}^{x_i}) - f(t_n, x_i, Y_{n, p}^i)]. \end{aligned} \quad (4.19)$$

Due to the Lipschitz continuity of f , it is easy to see that

$$|I_3| \leq C \Delta t |e_n^i|, \quad (4.20)$$

where C depends on the Lipschitz constant of f . We can split I_1 in the following way

$$\begin{aligned} I_1 &= \underbrace{\mathbb{E}_{t_n}^{x_i}[Y_{t_{n+1}}] - \mathbb{E}_{t_n, M_y}^{x_i}[Y_{t_{n+1}}]}_{I_{11}} + \underbrace{\mathbb{E}_{t_n, M_y}^{x_i}[Y_{t_{n+1}}] - \widehat{\mathbb{E}}_{t_n, M_y}^{x_i}[Y_{t_{n+1}}]}_{I_{12}} \\ &\quad + \underbrace{\widehat{\mathbb{E}}_{t_n, M_y}^{x_i}[Y_{t_{n+1}} - Y_{t_{n+1, p}}]}_{I_{13}} + \underbrace{\widehat{\mathbb{E}}_{t_n, M_y}^{x_i}[Y_{t_{n+1, p}} - Y_{n+1, p}]}_{I_{14}}, \end{aligned} \quad (4.21)$$

where $Y_{t_{n+1, p}}$ is the p -th order Lagrange polynomial interpolant of $Y_{t_{n+1}}$. By the definition of $\mathbb{E}_{t_n}^{x_i}[\cdot]$ and $\mathbb{E}_{t_n, M_y}^{x_i}[\cdot]$ in (3.8) and (3.9), we have

$$\begin{aligned} |I_{11}| &= \sum_{m=M_y+1}^{\infty} e^{-\lambda \Delta t} \frac{(\lambda \Delta t)^m}{m!} \int_E \cdots \int_E Y_{t_{n+1}} \left(x_i + \sum_{k=1}^m e_k \right) \left(\prod_{k=1}^m \rho(e_k) \right) de_1 \cdots de_m \\ &\leq C(\lambda \Delta t)^{M_y+1}, \end{aligned} \quad (4.22)$$

for $\lambda \Delta t < 1$, where C depends on the upper bounds of $|Y_{t_{n+1}}|$ and $|\rho|$. For the error I_{12} , we obtain

$$\begin{aligned} |I_{12}| &= \left| \mathbb{E}_{t_n, M_y}^{x_i}[Y_{t_{n+1}}] - \widehat{\mathbb{E}}_{t_n, M_y}^{x_i}[Y_{t_{n+1}}] \right| \\ &= \left| \sum_{m=M_y+1}^{\infty} e^{-\lambda \Delta t} \frac{(\lambda \Delta t)^m}{m!} \left\{ \int_E \cdots \int_E Y_{t_{n+1}} \left(x_i + \sum_{k=1}^m e_k \right) \left(\prod_{k=1}^m \rho(e_k) \right) de_1 \cdots de_m \right. \right. \\ &\quad \left. \left. - \sum_{q_1=1}^Q \cdots \sum_{q_m=1}^Q \left(\prod_{k=1}^m \tilde{w}_{q_k} \right) \cdot Y_{t_{n+1}} \left(x_i + \sum_{k=1}^m \tilde{a}_{q_k} \right) \right\} \right| \\ &\leq \sum_{m=M_y+1}^{\infty} e^{-\lambda \Delta t} \frac{(\lambda \Delta t)^m}{m!} \cdot (CQ^{-r}) \\ &\leq C \lambda \Delta t Q^{-r} \end{aligned} \quad (4.23)$$

for $\lambda \Delta t < 1$ where C depends on the upper bound of the derivatives of $Y_{t_{n+1}}$. For I_{13} , based on the fact that $Y_{t_{n+1}}(x_i) - Y_{t_{n+1, p}}(x_i) = 0$ for $x_i \in \mathcal{S}_h$ and the classic error

bound of Lagrange interpolation, we have

$$\begin{aligned}
|I_{13}| &= \left| \widehat{\mathbb{E}}_{t_n, M_y}^{x_i} [Y_{t_{n+1}} - Y_{t_{n+1}, p}] \right| \\
&\leq e^{-\lambda \Delta t} |Y_{t_{n+1}}(x_i) - Y_{t_{n+1}, p}(x_i)| + \sum_{m=1}^{M_y} e^{-\lambda \Delta t} \frac{(\lambda \Delta t)^m}{m!} \\
&\quad \cdot \left[\sum_{q_1=1}^Q \cdots \sum_{q_m=1}^Q \left(\prod_{k=1}^m \tilde{w}_{q_k} \right) \left| Y_{t_{n+1}} \left(x_i + \sum_{k=1}^m \tilde{a}_{q_k} \right) - Y_{t_{n+1}, p} \left(x_i + \sum_{k=1}^m \tilde{a}_{q_k} \right) \right| \right] \\
&\leq C(\Delta x)^{p+1} \cdot \sum_{m=1}^{M_1} e^{-\lambda \Delta t} \frac{(\lambda \Delta t)^m}{m!} \\
&\leq C\lambda \Delta t (\Delta x)^{p+1}.
\end{aligned} \tag{4.24}$$

Also, for I_{14} , we deduce that

$$\begin{aligned}
|I_{14}| &\leq \left| \widehat{\mathbb{E}}_{t_n, M_y}^{x_i} [Y_{t_{n+1}, p} - Y_{n+1, p}] \right| \\
&\leq e^{-\lambda \Delta t} |Y_{t_{n+1}, p}(x_i) - Y_{n+1, p}(x_i)| + \sum_{m=1}^{M_y} e^{-\lambda \Delta t} \frac{(\lambda \Delta t)^m}{m!} \cdot \Lambda_p^{\Delta x} \max_{i \in \mathbb{Z}} |e_{n+1}^i| \\
&\leq (1 + C\Delta t \Lambda_p^{\Delta x}) \max_{i \in \mathbb{Z}} |e_{n+1}^i|,
\end{aligned} \tag{4.25}$$

where $\Lambda_p^{\Delta x}$ is the Lebesgue constant of the p -th order Lagrange interpolant under the mesh \mathcal{S}_h .

Combining $I_{11}, I_{12}, I_{13}, I_{14}$, we obtain the following upper bound of $|I_1|$, i.e.,

$$|I_1| \leq C [(\lambda \Delta t)^{M_y+1} + \lambda \Delta t Q^{-r} + \lambda \Delta t (\Delta x)^{p+1}] + (1 + C\Delta t \Lambda_p^{\Delta x}) \max_{i \in \mathbb{Z}} |e_{n+1}^i|. \tag{4.26}$$

Similarly, we can obtain the following upper bound of $|I_2|$, i.e.,

$$|I_2| \leq C\Delta t \left[(\lambda \Delta t)^{M_f+1} + \lambda \Delta t Q^{-r} + \lambda \Delta t (\Delta x)^{p+1} + (1 + C\Delta t \Lambda_p^{\Delta x}) \max_{i \in \mathbb{Z}} |e_{n+1}^i| \right]. \tag{4.27}$$

Substituting (4.26), (4.27) and (4.20) into (4.18), we have

$$\begin{aligned}
(1 - C\Delta t) \max_{i \in \mathbb{Z}} |e_n^i| &\leq (1 + C\Delta t) \max_{i \in \mathbb{Z}} |e_{n+1}^i| \\
&\quad + C [(\lambda \Delta t)^{M_y+1} + \Delta t (\lambda \Delta t)^{M_f+1}] + \lambda \Delta t Q^{-r} + C\lambda \Delta t (\Delta x)^{p+1} + C(\Delta t)^3.
\end{aligned} \tag{4.28}$$

Therefore, for sufficiently small Δt , we obtain, by induction,

$$\max_{i \in \mathbb{Z}} |e_n^i| \leq C [(\Delta t)^2 + (\lambda \Delta t)^{M_y} + (\lambda \Delta t)^{M_f+1} + Q^{-r} + (\Delta x)^{p+1}]. \tag{4.29}$$

□

5. Numerical examples. In this section, we report on the results of three one-dimensional numerical examples, that illustrate the accuracy and efficiency of the proposed schemes (3.10). We take uniform partitions in both temporal and spatial domains with time and space step sizes Δt and Δx , respectively. The number of

time steps is denoted by N , which is given by $N = \frac{T}{\Delta t}$, with T is the terminal time. For the sake of illustration, we only solve the nonlocal problems on bounded spatial domains. Lagrange interpolation and tensor-product quadrature rules are used to approximate $\mathbb{E}_{t_n}^{x_i}[\cdot]$, where the one-dimensional quadrature rule for each example is chosen according to the property of the kernel $\gamma(e)$, the initial condition u_0 and the forcing term f . Our main goal is to test the accuracy and convergence rate of the proposed scheme (3.10) with respect to Δt and Δx . To this end, according to the analysis in Theorem 3, we always set the number of quadrature points to be sufficiently large so that the error contributed by the use of quadrature rules is too small to affect the convergence rate.

5.1. Symmetric kernel. First we consider the following nonlocal diffusion problem in $[0, T]$:

$$\begin{cases} \frac{\partial u}{\partial t} - \frac{1}{\delta^3} \int_{-\delta}^{\delta} (u(t, x+e) - u(t, x)) de = g(t, x), & t > 0, \\ u(0, x) = \varphi(x), \end{cases} \quad (5.1)$$

where $\delta > 0$ which corresponds to the symmetric kernel

$$\gamma(e) = \begin{cases} \frac{1}{\delta^3}, & \text{for } e \in [-\delta, \delta], \\ 0, & \text{for } e \notin [-\delta, \delta]. \end{cases} \quad (5.2)$$

We choose the exact solution to be

$$u(t, x) = (-x^3 + x^2) \exp\left(-\frac{t}{10}\right),$$

so that the forcing term g is given by

$$g(t, x) = -\frac{u(t, x)}{10} + \left(2x - \frac{2}{3}\right) \exp\left(-\frac{t}{10}\right),$$

and the initial condition $\varphi(x)$ can be determined from u accordingly. After converting the problem (5.1) to a BSDE of the form in (2.12), we have

$$\lambda = \frac{2}{\delta^2}, \quad \rho(e) = \frac{1}{2\delta} \mathcal{I}_{[-\delta, \delta]}(e), \quad b = 0, \quad f(t, \cdot, \cdot) = g(T - t, \cdot, \cdot),$$

where $\mathcal{I}_{[-\delta, \delta]}(e)$ is a characteristic function of the interval $[-\delta, \delta]$.

We set the terminal time $T = 1$ and solve (5.1) on the spatial domain $[0, 1] \in \mathbb{R}$. Since the density function $\rho(e)$ is uniform with the support $[-\delta, \delta]$, we use tensor-product Gauss-Legendre quadrature rules, with $Q = 16$, to approximate the integrals involved in $\mathbb{E}_{t_n}^{x_i}[\cdot]$. First, we test the convergence rate with respect to Δt . To this end, we set $N_x = 65$ and use piecewise cubic Lagrange interpolation ($p = 3$) to construct $Y_{n,p}$ for $n = 0, \dots, N - 1$, so that the time discretization error dominates the total error. For $\delta = 1$, the number N of time steps is set to 4, 8, 16, 32, 64, respectively. The numerical results are shown in Table 5.1. As expected, the convergence rate with respect to Δt depends not only on the value of θ , but also the number of jumps retained in constructing $\mathbb{E}_{t_n, M_y}^{x_i}[\cdot]$ and $\mathbb{E}_{t_n, M_f}^{x_i}[\cdot]$. For example, when $M_y = M_f = 0$, i.e., no jump is included in $\mathbb{E}_{t_n, M_y}^{x_i}[\cdot]$ and $\mathbb{E}_{t_n, M_f}^{x_i}[\cdot]$, the scheme (3.10) fails to converge. In general, in order to achieve a k -th order convergence ($k = 1, 2$), M_y and M_f must satisfy $M_y \geq k$ and $M_f \geq k - 1$.

Table 5.1: Errors and convergence rates with respect to Δt in Example 5.1 where $T = 1$, $\delta = 1$, $N_x = 65$, $p = 3$.

	$\ Y_{t_0} - Y_{0,p}\ _\infty$					
	$N = 4$	$N = 8$	$N = 16$	$N = 32$	$N = 64$	CR
$\theta = 0, M_y = 0, M_f = 0$	1.932E-1	1.978E-1	2.000E-1	2.012E-1	2.017E-1	-0.015
$\theta = 0, M_y = 1, M_f = 0$	7.958E-2	3.435E-2	1.572E-2	7.487E-3	3.649E-3	1.109
$\theta = 0, M_y = 2, M_f = 1$	3.673E-2	1.849E-2	9.279E-3	4.675E-3	2.326E-3	0.995
$\theta = 1, M_y = 0, M_f = 0$	2.111E-1	2.067E-1	2.045E-1	2.034E-1	2.028E-1	0.014
$\theta = 1, M_y = 1, M_f = 0$	4.891E-2	2.581E-2	1.328E-2	6.736E-3	3.393E-3	0.964
$\theta = 1, M_y = 2, M_f = 1$	6.170E-2	3.007E-2	1.468E-2	7.231E-3	3.585E-3	1.027
$\theta = \frac{1}{2}, M_y = 0, M_f = 0$	2.030E-1	2.025E-1	2.023E-1	2.023E-1	2.023E-1	0.001
$\theta = \frac{1}{2}, M_y = 1, M_f = 0$	2.623E-2	1.326E-2	6.666E-3	3.342E-3	1.674E-3	0.993
$\theta = \frac{1}{2}, M_y = 2, M_f = 1$	3.207E-3	8.258E-4	2.094E-4	5.271E-5	1.322E-5	1.981
$\theta = \frac{1}{2}, M_y = 3, M_f = 2$	3.330E-3	8.632E-4	2.196E-4	5.538E-5	1.391E-5	1.977

We observe in Theorem 3 that the errors with respect to Δt from $\mathbb{E}_{t_n, M_y}^{x_i}[\cdot]$ and $\mathbb{E}_{t_n, M_f}^{x_i}[\cdot]$ are of order $(\lambda \Delta t)^{M_y}$ and $(\lambda \Delta t)^{M_f+1}$, respectively. If the intensity λ is large, i.e., the horizon δ is small, the value of δ will affect the error decay in the pre-asymptotic region and change the total error up to a constant. Such phenomenon is investigated by setting $\delta = 0.3$ and 0.4 , and $\theta = 1$ and $\frac{1}{2}$. The results are shown in Figure 5.1. In Figure 5.1(a) with $\theta = 1$, in the case that $\delta = 0.3$ ($\lambda \approx 22$), $M_y = 1$ and $M_f = 0$, the error is actually of order $\mathcal{O}(\lambda \Delta t)$. When one more jump is included, i.e., $M_y = 2$ and $M_f = 1$, we observe that the convergence rate remains the same but the total error is significantly reduced, since the error contributed by $\mathbb{E}_{t_n, M_y}^{x_i}[\cdot]$ and $\mathbb{E}_{t_n, M_f}^{x_i}[\cdot]$ is reduced to $\mathcal{O}((\lambda \Delta t)^2)$.

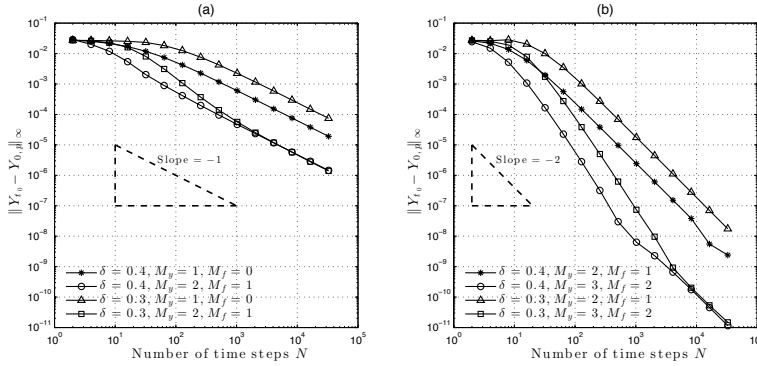


Fig. 5.1: Error decays for (a) $\theta = 1$ and (b) $\theta = \frac{1}{2}$ in Example 5.1.

Next, we test the convergence rate with respect to Δx by setting $\delta = 1$, $T = 1$, $\theta = \frac{1}{2}$, $N = 1024$, $M_y = 3$ and $M_f = 2$. We divide the spatial interval $[0, 1]$ into

N_x elements with $\Delta x = 2^{-3}, 2^{-4}, 2^{-5}, 2^{-6}, 2^{-7}$. The error is measured in L^∞ and L^2 norms. In Table 5.2, we can see that the numerical results verify the theoretical analysis in §4.

Table 5.2: Errors and convergence rates with respect to Δx in Example 5.1 where $\delta = 1$, $T = 1$, $\theta = \frac{1}{2}$, $N = 1024$, $M_y = 3$ and $M_f = 2$.

	Linear interpolation		Quadratic interpolation	
Δx	$\ Y_{t_0} - Y_{0,1}\ _{L^2}$	$\ Y_{t_0} - Y_{0,1}\ _{L^\infty}$	$\ Y_{t_0} - Y_{0,2}\ _{L^2}$	$\ Y_{t_0} - Y_{0,2}\ _{L^\infty}$
2^{-3}	1.227E-03	2.816E-03	1.862E-04	2.969E-04
2^{-4}	3.266E-04	7.356E-04	2.278E-05	3.881E-05
2^{-5}	6.586E-05	1.772E-04	2.792E-06	4.661E-06
2^{-6}	1.651E-05	4.481E-05	3.491E-07	6.252E-07
2^{-7}	4.167E-06	1.103E-05	4.738E-08	9.188E-08
CR	2.071	2.001	2.991	2.927

5.2. Singular kernel. We consider the following nonlocal diffusion problem in $[0, T]$

$$\begin{cases} \frac{\partial u}{\partial t} - \frac{1}{\delta^2 \sqrt{\delta}} \int_{-\delta}^{\delta} \frac{u(t, x+e) - u(t, x)}{\sqrt{|e|}} de = g(t, x), & t > 0, \\ u(0, x) = \varphi(x), \end{cases} \quad (5.3)$$

where $\delta > 0$ which corresponds to the singular kernel $\gamma(e)$

$$\gamma(e) = \begin{cases} \frac{1}{\delta^2 \sqrt{\delta|e|}}, & \text{for } e \in [-\delta, \delta], \\ 0, & \text{for } e \notin [-\delta, \delta]. \end{cases} \quad (5.4)$$

We choose the exact solution to be

$$u(t, x) = (x + t)^2,$$

so that the forcing term g is given by

$$g(t, x) = 2(x + t) - \frac{4}{5}$$

and the initial condition $\varphi(x)$ can be determined from u accordingly. After converting the problem (5.1) to a BSDE of the form in (2.12), we have

$$\lambda = \frac{4}{\delta^2}, \quad \rho(e) = \frac{1}{4\sqrt{\delta|e|}} \mathcal{I}_{[-\delta, \delta]}(e), \quad b = 0, \quad f(t, \cdot, \cdot) = g(T - t, \cdot, \cdot),$$

where $\mathcal{I}_{[-\delta, \delta]}(e)$ is a characteristic function of the interval $[-\delta, \delta]$.

We set the terminal time $T = 0.25$ and solve the equation (5.3) on the spatial domain $[0, 1] \in \mathbb{R}$. It is observed that the density function $\rho(e)$ is singular at $e = 0$, but it is still integrable in the interval $[-\delta, \delta]$. Recalling that $1/\sqrt{|e|}$ is a special case

of the kernel of the Gauss-Jacobi quadrature rule, i.e. $\int_a^b \psi(x)(b-x)^\alpha(x-a)^\beta dx$ with $\alpha, \beta \in [-1, 1]$, where the kernel is given by $(b-x)^\alpha(x-a)^\beta$. Thus, the integrals involved in $\mathbb{E}_{t_n, M_y}^{x_i}[\cdot]$ and $\mathbb{E}_{t_n, M_f}^{x_i}[\cdot]$ can be accurately approximated by setting $a = 0$, $b = \delta$, $\alpha = 0$ and $\beta = -\frac{1}{2}$. In this example, we use 16-point Gauss-Jacobi rule such that the quadrature error can be ignored. We test the convergence rates with respect to Δt by setting $\delta = 1$, $N_x = 65$ and $p = 3$, and the results are given in Table 5.3. As expected, due to the use of Gauss-Jacobi rule, the temporal truncation error dominates the total error and the theoretical result in Theorem 3 has been verified as well. The convergence with respect to Δx is also tested by setting $p = 1$, $T = 0.25$, $\theta = \frac{1}{2}$, $N = 1024$, $M_y = 3$ and $M_f = 2$. Table 5.4 shows the results in two cases where $\delta = 1$ and 0.1 , respectively. We see that, for different horizon values, the spatial discretization error decays at the same rate verifying the theoretical error estimates in §4.

Table 5.3: Errors and convergence rates with respect to Δt in Example 5.2 where $T = 0.25$, $\delta = 1$, $N_x = 65$, $p = 3$.

$\ Y_{t_0} - Y_{0,p}\ _\infty$						
	$N = 4$	$N = 8$	$N = 16$	$N = 32$	$N = 64$	CR
$\theta = 0, M_y = 0, M_f = 0$	9.732E-1	9.554E-1	9.461E-1	9.413E-1	9.390E-1	0.013
$\theta = 0, M_y = 1, M_f = 0$	2.099E-1	1.108E-1	5.698E-2	2.890E-2	1.456E-2	0.964
$\theta = 0, M_y = 2, M_f = 1$	8.618E-2	4.131E-2	2.013E-2	9.925E-3	4.926E-3	1.032
$\theta = 1, M_y = 0, M_f = 0$	8.958E-1	9.167E-1	9.267E-1	9.317E-1	9.341E-1	-0.014
$\theta = 1, M_y = 1, M_f = 0$	1.239E-1	6.669E-2	3.461E-2	1.763E-2	8.900E-3	0.952
$\theta = 1, M_y = 2, M_f = 1$	8.485E-2	5.274E-2	2.959E-2	1.577E-2	8.149E-3	0.950
$\theta = \frac{1}{2}, M_y = 0, M_f = 0$	9.345E-1	9.360E-1	9.364E-1	9.365E-1	9.365E-1	-0.001
$\theta = \frac{1}{2}, M_y = 1, M_f = 0$	1.669E-1	8.874E-2	4.579E-2	2.327E-2	1.173E-2	0.959
$\theta = \frac{1}{2}, M_y = 2, M_f = 1$	1.634E-2	4.386E-3	1.136E-3	2.889E-4	7.286E-5	1.954
$\theta = \frac{1}{2}, M_y = 3, M_f = 2$	4.476E-3	1.079E-3	2.634E-4	6.498E-5	1.613E-5	2.029

Table 5.4: Errors and convergence rates with respect to Δx in Example 5.2 where $p = 1$, $T = 0.25$, $\theta = 1/2$, $N = 1024$, $M_y = 3$ and $M_f = 2$.

Δx	$\delta = 1$		$\delta = 0.1$	
	$\ Y_{t_0} - Y_{0,p}\ _{L^2}$	$\ Y_{t_0} - Y_{0,p}\ _{L^\infty}$	$\ Y_{t_0} - Y_{0,p}\ _{L^2}$	$\ Y_{t_0} - Y_{0,p}\ _{L^\infty}$
2^{-3}	3.912E-03	5.232E-03	7.892E-03	1.046E-02
2^{-4}	9.934E-04	1.318E-03	2.030E-03	2.618E-03
2^{-5}	2.647E-04	3.482E-04	4.893E-04	5.787E-04
2^{-6}	6.062E-05	8.057E-05	1.269E-04	1.180E-04
2^{-7}	1.318E-05	1.862E-05	1.498E-05	2.541E-05
CR	2.046	2.030	2.208	2.184

5.3. Non-symmetric kernel and discontinuous solution. We consider the following nonlocal diffusion problem in $[0, T]$,

$$\begin{cases} \frac{\partial u}{\partial t} - \int_{-\delta}^{2\delta} [u(t, x+e) - u(t, x)] de = g(t, x), & t > 0, \\ u(0, x) = \varphi(x), \end{cases} \quad (5.5)$$

where $\delta > 0$ for which we have the non-symmetric kernel

$$\gamma(e) = \begin{cases} 1, & \text{if } e \in [-\delta, 2\delta], \\ 0, & \text{if } e \notin [-\delta, 2\delta]. \end{cases} \quad (5.6)$$

We choose the exact solution

$$u(t, x) = \begin{cases} x \sin(t), & \text{if } x < \frac{1}{2}, \\ x^2 \sin(t), & \text{if } x \geq \frac{1}{2} \end{cases} \quad (5.7)$$

so that the forcing term g is given by

$$g(t, x) = \begin{cases} \sin(t) \left[-\frac{(x+2\delta)^2}{2} + \frac{(x-\delta)^2}{2} + 3\delta x \right] + x \cos(t), & x < \frac{1}{2} - 2\delta, \\ \sin(t) \left[-\frac{1}{12} - \frac{(x+2\delta)^3}{3} + \frac{(x-\delta)^2}{2} + 3\delta x \right] + x \cos(t), & \frac{1}{2} - 2\delta \leq x < \frac{1}{2}, \\ \sin(t) \left[-\frac{1}{12} - \frac{(x+2\delta)^3}{3} + \frac{(x-\delta)^2}{2} + 3\delta x^2 \right] + x^2 \cos(t), & \frac{1}{2} \leq x < \frac{1}{2} + \delta, \\ \sin(t) \left[-\frac{(x+2\delta)^3}{3} + \frac{(x-\delta)^3}{3} + 3\delta x^2 \right] + x^2 \cos(t), & x \geq \frac{1}{2} + \delta \end{cases} \quad (5.8)$$

and the initial condition $\varphi(x)$ can be determined from u accordingly. After converting the problem (5.5) to a BSDE of the form in (2.12), we have

$$\lambda = 3\delta, \quad \rho(e) = \frac{1}{3\delta} \mathcal{I}_{[-\delta, 2\delta]}(e), \quad b = \frac{3}{2} \delta^2, \quad f(t, \cdot, \cdot) = g(T-t, \cdot, \cdot).$$

Note that the kernel $\gamma(e)$ is non-symmetric so the drift coefficient b is non-zero. However, as shown in (2.11), non-zero b is equal to the compensator of the underlying compound Poisson process, such that it cancels out the compensator in the forward SDE in (2.12), so that it does not appear in the numerical scheme (3.10). The most challenging aspect of this problem is the discontinuities in u and g , which deteriorate the accuracy of the quadrature rule and the polynomial interpolation. Here, the trapezoidal quadrature rule so that the integral of the interpolating polynomials are evaluated exactly. To study the convergence of the scheme (3.10) with respect to Δx , we set $T = 0.5$, $\theta = \frac{1}{2}$ and $N = 512$. First, we solve (5.5) on a uniform spatial grid with $\Delta x = 2^{-3}, 2^{-4}, 2^{-5}, 2^{-6}, 2^{-7}$. The errors and convergence rates are given in Table 5.5. As we expected, due to the discontinuity, the L^∞ error does not converge and the L^2 error converges with $\mathcal{O}((\Delta x)^{\frac{1}{2}})$.

In order to improve the convergence rate with respect to the L^2 norm, one strategy is to utilize adaptive grids which are capable of automatically refining the spatial grid

Table 5.5: Errors and convergence rates with respect to Δx in Example 5.3 for $p = 1$, $T = 0.5$, $\theta = \frac{1}{2}$, $N = 512$.

	$\delta = 1$		$\delta = 0.1$	
Δx	$\ Y_{t_0} - Y_{0,p}\ _{L^2}$	$\ Y_{t_0} - Y_{0,p}\ _{L^\infty}$	$\ Y_{t_0} - Y_{0,p}\ _{L^2}$	$\ Y_{t_0} - Y_{0,p}\ _{L^\infty}$
2^{-3}	3.001E-02	1.323E-01	2.503E-02	1.215E-01
2^{-4}	2.287E-02	1.317E-01	1.751E-02	1.207E-01
2^{-5}	1.467E-02	1.257E-01	1.231E-02	1.201E-01
2^{-6}	1.107E-02	1.254E-01	8.676E-03	1.197E-01
2^{-7}	7.045E-03	1.221E-01	6.126E-03	1.192E-01
CR	0.523	0.030	0.507	0.007

around the discontinuity. Here, we apply the one-dimensional adaptive hierarchical finite element method [12, 19]. This approach can be seamlessly integrated into the sparse grid framework to alleviate the curse of dimensionality when solving high-dimensional nonlocal diffusion problems. Figure 5.2(a) shows the solution surface $u(t, x)$ we recovered in order to approximate $u(0.5, x)$ within the spatial domain $[0, 1]$; and Figure 5.2(b) illustrates the exact solution $u(0.5, x)$ and its approximation using 33 spatial grid points for which the L^2 error is only 9.745E-04. Further illustration is shown in Figure 5.3(a) and Figure 5.3(b) where, for the scheme (3.10), the error decay vs. the number of interpolating points and vs. the grid size Δx , respectively, of using uniform and adaptive grids are plotted. For the adaptive case, Δx is the starting grid for the refinement process. We clearly see the half-order and the optimal second-order convergence rates for the uniform and adaptive grid cases, respectively. We also study the convergence with respect to Δt for $\theta = 1$ with the error tolerance for the adaptive grid is set to 0.01, 0.005, 0.001 and 0.0005. In Figure 5.3(c), it can be seen that the desired convergence rate with respect to Δt is achieved if the error is smaller than the tolerance; otherwise, the spatial discretization error will dominate and the total error will not decay as Δt decreases.

Of course, no amount of refinement, whether adaptive or not, will reduce the L^∞ norm error, i.e., it will remain $\mathcal{O}(1)$. However, if we omit the single element containing the discontinuity, we see the following from Figure 5.3(d). For uniform grid refinement, the error remains $\mathcal{O}(1)$ due to pollution of the error into neighboring elements which are all of size Δx . However, for adaptive grid refinement, the L^∞ error is close to $\mathcal{O}(h^2)$ because pollution is to neighboring elements of very small size.

6. Concluding remarks and future work. In this work, we propose a novel stochastic numerical scheme for linear nonlocal diffusion problems based on the relationship between the PIDEs and a certain class of backward stochastic differential equations (BSDEs) with jumps. As discussed in §1, this effort focuses on solving low-dimensional nonlocal diffusion problems with high-order accuracy, where the potential applications will be non-Darcy flow in groundwater, or plasma transport in magnetically confined fields. However, our approach can be extended to solve moderately high-dimensional problems by incorporating our previous work on sparse-grid approximation [19, 12]. Compared to standard finite element and collocation approaches for approximating linear nonlocal diffusion equations with integrable kernels, our method completely avoids the solution of dense linear systems. Moreover, the ability to utilize

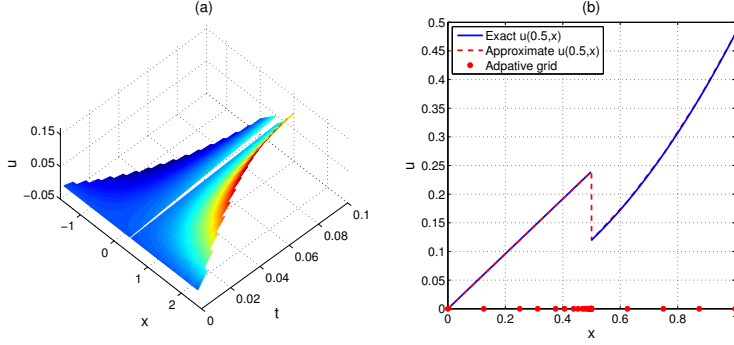


Fig. 5.2: (a) The surface of $u(t, x)$; (b) The exact solution $u(0.5, x)$ (solid line) and its approximation (dashed line) using 33 grid points (red dots).

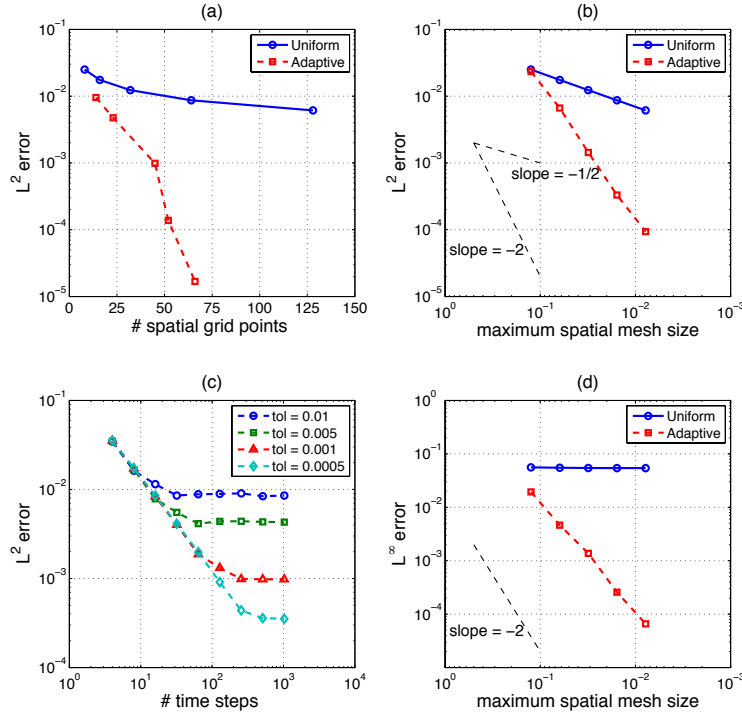


Fig. 5.3: Convergence of the L^2 norm of the error with respect to (a) the number of spatial grid points and (b) the spatial grid size Δx using uniform and adaptive grids. (c) Convergence with respect to Δt for various tolerances of the adaptive spatial grid. (d) Convergence of the L^∞ norm of the error (with the element containing the discontinuity ignored) with respect to Δx using uniform and adaptive grids.

high-order temporal and spatial discretization schemes, as well as efficient adaptive approximation, and the potential of massively parallel implementation, make our technique highly advantageous. These assets have been verified by both theoretical

analysis and numerical experiments.

Our future efforts will focus on extending the proposed numerical schemes to the case of non-integrable kernels, e.g., fractional Laplacians, where the underlying Lévy processes cannot be described by compound Poisson processes. Moreover, the well-posedness of BSDEs on bounded domains, with volume constraints, is also critical for studying nonlocal diffusion equations on bounded domains.

REFERENCES

- [1] G. BARLES, R. BUCKDAHN, AND E. PARDOUX, *Backward stochastic differential equations and integral-partial differential equations*, Stochastics Stochastics Rep., 60 (1997), pp. 57–83.
- [2] B. BOUCHARD AND R. ELIE, *Discrete-time approximation of decoupled Forward–Backward SDE with jumps*, Stochastic Process. Appl., 118 (2008), pp. 53–75.
- [3] B. BOUCHARD, R. ELIE, AND N. TOUZI, *Discrete-time approximation of BSDEs and probabilistic schemes for fully nonlinear PDEs*, in Advanced financial modelling, Walter de Gruyter, Berlin, 2009, pp. 91–124.
- [4] H.-J. BUNGARTZ AND M. GRIEBEL, *Sparse grids*, Acta Numerica, 13 (2004), pp. 1–123.
- [5] N. BURCH AND R. B. LEHOUCQ, *Continuous-time random walks on bounded domains*, Physical Review E, 83 (2011), p. 12105.
- [6] X. CHEN AND M. GUNZBURGER, *Continuous and discontinuous finite element methods for a peridynamics model of mechanics*, Computer Methods in Applied Mechanics and Engineering, 200 (2011), pp. 1237–1250.
- [7] L. DELONG, *Backward Stochastic Differential Equations with Jumps and Their Actuarial and Financial Applications*, BSDEs with Jumps, Springer Science & Business, June 2013.
- [8] Q. DU, M. GUNZBURGER, R. B. LEHOUCQ, AND K. ZHOU, *Analysis and approximation of nonlocal diffusion problems with volume constraints*, SIAM Rev., 54 (2012), pp. 667–696.
- [9] ———, *A nonlocal vector calculus, nonlocal volume-constrained problems, and nonlocal balance laws*, Math. Models Methods Appl. Sci., 23 (2013), pp. 493–540.
- [10] Q. DU, Z. HUANG, AND R. B. LEHOUCQ, *Nonlocal convection-diffusion volume-constrained problems and jump processes*, Discrete and Continuous Dynamical Systems - Series B, 19 (2014), pp. 373–389.
- [11] Q. DU, L. JU, L. TIAN, AND K. ZHOU, *A posteriori error analysis of finite element method for linear nonlocal diffusion and peridynamic models*, Math. Comp., 82 (2013), pp. 1889–1922.
- [12] M. D. GUNZBURGER, C. G. WEBSTER, AND G. ZHANG, *Stochastic finite element methods for partial differential equations with random input data*, Acta Numerica, 23 (2014), pp. 521–650.
- [13] F. B. HANSON, *Applied stochastic processes and control for Jump-diffusions: modeling, analysis, and computation*. SIAM, Philadelphia, 2007.
- [14] A. LEJAY, E. MORDECKI, AND S. TORRES, *Numerical approximation of Backward Stochastic Differential Equations with Jumps*, (2007).
- [15] R. METZLER AND J. KLAFTER, *The random walk’s guide to anomalous diffusion: a fractional dynamics approach*, Physics Reports, 339 (2000), pp. 1–77.
- [16] E. PARDOUX AND S. PENG, *Backward stochastic differential equations and quasilinear parabolic partial differential equations*, in Stochastic Partial Differential Equations and Their Applications, Springer Berlin Heidelberg, Berlin/Heidelberg, Jan. 1992, pp. 200–217.
- [17] E. PARDOUX AND S. G. PENG, *Adapted solution of a backward stochastic differential equation*, Systems & Control Letters, 14 (1990), pp. 55–61.
- [18] E. PLATEN AND N. BRUTI-LIBERATI, *Numerical Solution of Stochastic Differential Equations with Jumps in Finance*, vol. 64 of Stochastic Modelling and Applied Probability, Springer Berlin Heidelberg, Berlin, Heidelberg, 2010.
- [19] G. ZHANG, M. GUNZBURGER, AND W. ZHAO, *A sparse-grid method for multi-dimensional backward stochastic differential equations*, Journal of Computational Mathematics, 31 (2013), pp. 221–248.
- [20] W. ZHAO, L. CHEN, AND S. PENG, *A New Kind of Accurate Numerical Method for Backward Stochastic Differential Equations*, SIAM J. Sci. Comput., 28 (2006), pp. 1563–1581.
- [21] W. ZHAO, Y. LI, AND G. ZHANG, *A generalized θ -scheme for solving backward stochastic differential equations*, Discrete Contin. Dyn. Syst. Ser. B, 17 (2012), pp. 1585–1603.
- [22] W. ZHAO, G. ZHANG, AND L. JU, *A stable multistep scheme for solving backward stochastic differential equations*, SIAM J. Numer. Anal., 48 (2010), pp. 1369–1394.